Einführung

in die

Statistik

A. B. Zeidler

HfWU Nürtingen

SS 2008 (99 Seiten)

Inhaltsverzeichnis

Kapitel 1

Einführung

Was ist Statistik?
Unter einer *Statistik* versteht man im allgemeinen Sprachgebrauch eine geordnete Zusammenstellung von empirisch erhobenen Daten. Andererseits bezeichnet das Wort Statistik aber auch die mathematischen Verfahren solche Daten aufzubereiten.

Zunächst müssen die Daten aber erst einmal gesammelt werden. Im günstigsten Fall kann man alle benötigten Daten erheben - etwa bei einer Volkszählung, in der wirklich jeder Bürger erfasst wird. In diesem Fall spricht man von einer **Vollerhebung**. Oft ist es aber zu aufwendig oder gar nicht möglich wirklich alle Daten zu erheben. Will man z.B. wissen, wieviel Prozent aller produzierten Feuerwerkskörper tatsächlich explodieren, macht es wenig Sinn, alle zu testen. Wertet man also nur eine *Stichprobe* aus, spricht man von einer **Teilerhebung**.

Die *beschreibende Statistik* befasst sich mit der Auswertung vollständiger Datensätze, ist also die Statistik der Vollerhebungen. Es geht dabei nur um die Aufbereitung der Daten. In der *schließenden Statistik* wertet man nur Stichproben aus, sie ist also die Statistik der Teilerhebungen. Man versucht also von der Stichprobe auf die Gesamtheit zu schließen. Solche Ergebnisse haben naturgemäss nur eine gewisse Wahrscheinlichkeit zuzutreffen.

Das Ziel der Statistik:
Alle (guten) Statistiken verfolgen ein und dasselbe Ziel: sie sollen ermöglichen eine angemessene, qualifizierte Entscheidung zu treffen. Beispiele:

Statistik		Entscheidung
Bevölkerungsstatistik	$\rightarrow$	Rentenversicherungsbeiträge
volkswirtschaftliche S.	$\rightarrow$	Leitzins, Steuerpolitik
Unternehmensstatistiken	$\rightarrow$	Investitionen, Produktion, Marketing

Im Idealfall wird also erst die Statistik erhoben und anhand derer bildet man sich eine Meinung. Leider wird dies in der Praxis (insbesondere in der Politik) gerne auf den Kopf gestellt: erst ist die Meinung und dann werden Statistiken erhoben oder frisiert, die diese Meinung stützen sollen.

Versucht man bei der Entscheidungsfindung alle Einzeldaten zu berücksichtigen, sieht man leicht den Wald vor lauter Bäumen nicht mehr. Die Datenflut muss erst komprimiert werden, wobei zwingend Informationen verloren gehen (müssen). Die Statistik hat damit zwei konkurrierende Ziele:

1. möglichst hohe Kompression der Daten, bei
2. möglichst geringem Informationsverlust.

Die Statistik hat also die Aufgabe diesen Spagat so gut wie möglich zu meistern. Und um dies zu tun, muss man genau analysieren, wie viel Information verloren ging, bzw. erhalten blieb. An dieser Stelle wird klar, warum die Statistik zwingend mathematische Methoden verwenden muss.

Begriffsbildung:

Bezeichnung	Symbol	Bedeutung
Merkmalsträger	t	= die Objekte, deren
= stat. Einheiten		Daten betrachtet werden
statistische Masse	Ω	= Gesamtheit aller Merkmalsträger
		= die Menge der statistischen Einheiten
Merkmal	s	= die betrachteten Daten
Ausprägung	x_t	= der tatsächliche (Zahlen)wert des
		Merkmals beim Merkmalsträger t
Merkmalsraum	S	= Bereich aller denkbaren Ausprägungen
		= Menge, die die Ausprägungen enthält

Wir werden die Merkmalsträger im folgenden stets mit 1, 2 bis n durchnummerieren. Es ist damit also $\Omega = 1 \dots n$ und n gibt die Zahl der Merkmalsträger an. Wegen $t \in \Omega$ ist t dann also eine Zahl von 1 bis n. Ist der Merkmalsraum endlich, verwenden wir $S = \{ s_1, s_2, \dots, s_m \}$. Die s_i sind also die *verschiedenen*, *möglichen* Ausprägungen und m ist die Zahl, wieviele es davon gibt. Die x_t hingegen sind die tatsächlich vorkommenden Ausprägungen und daher, muss immer $x_t \in S$ sein. Man beachte, dass die x_t aber *nicht* verschieden sein müssen.

Beispiel: Zum Abschluss der Vorlesung *Statistik* schreiben die Studenten eine Klausur, die von den Dozenten korrigiert und mit Schulnoten bewertet wird. In diesem Fall sind die Merkmalsträger die Studenten, das Merkmal ist die Klausurnote und der Merkmalsraum ist die Notenskala von 1 bis 6.

Beispiel: Der Bekanntheitsgrad der Regierungsmitglieder soll durch eine Umfrage unter 2000 Bürgern gemessen werden. Dem Befragten wird der Name des Politikers genannt, er soll dessen Funktion nennen (z.B. Müntefering ist Viezekanzler). Der Bekanntheitsgrad soll der Anteil richtiger Antworten sein. Dann sind die Merkmalsträger die Regierungsmitglieder (nicht die Befragten!), das Merkmal ist der Bekanntheitsgrad und der Merkmalsraum sind die Prozentzahlen von 0% bis 100%.

Systematik:

Es gibt viele verschiedene Arten von Merkmalen. Grundsätzlich unterscheidet man Merkmale erst einmal nach 3 Kriterien: der Erhebungsart, der Skalierung und der Vergleichbarkeit. Jedes dieser Kriterien hat selbst wieder verschiedene mögliche Ausprägungen:

1. **Erhebungsart:** Man unterscheidet *stacks* (Bestandsma/ssen), die zu einem bestimmten Zeitpunkt erfasst werden und *flows* (Bewegungsma/ssen), die über einen Zeitraum erfasst werden. Beispiele für stacks sind das Alter, die Zahl der Mitarbeiter oder Kontostände. Beispiele für flows sind die Zahl der Einstellungen oder Unternehmensgründungen (in einem gegbenen Zeitraum).

2. **Skalierung:** Man unterscheidet *diskrete* und *kontinierliche* Merkmale. Ein Merkmal heißt diskret, wenn die Ausprägung des Merkmals ist in (abzählbar viele) Stufen unterteilt ist. Jede Ausprägung hat also einen offensichtlichen Nachfolger. Beipiele sind Jahreszahlen, Handelsklassen und Kontostände (in Cent). Ein Merkmal heißt kontinuierlich, wenn die Ausprägungen des Merkmals (unendlich viele) Zwischenschritte erlauben. Beispiele sind Entfernungen, Arbeitszeiten und Umrechnungskurse (z.B. von Dollar in Euro).

3. **Vergleichbarkeit:** Man unterscheidet *qualitative* Merkmale, bei denen man nur sagen kann ob zwei Ausprägungen gleich oder verschieden sind, und *komparative* Merkmale, bei denen die Merkmale eine natürliche Ordnung besizen. Beispiele für rein qualitative Merkmale sind Farbe, Geschlecht, Name, PLZ oder Beruf. Rein komparative Merkmale sind Handelsklasse, Berufsbildung, Körbchengröße oder Schulnote. Sind die Ausprägungen sogar Zahlenwerte (also $S \subseteq \mathbb{R}$) so sprechen wir von einem *quantitativen* Merkmal, z.B. Gewinn oder Kontostand. In diesem Fall kann man auch den Abstand zwischen zwei Werten berechnen. Bei *positiven* Merkmalen (also $S \subseteq \mathbb{R}^+$) gibt es sogar einen absoluten Bezugspunkt 0, z.B. bei Gehalt, Preis oder Alter. In diesem Fall ist es auch sinnvoll Verhältnisse zu bilden.

Kapitel 2

Mengenlehre

Mathematik ist eine abstrakte Wissenschaft, die sich mit beliebigen Objekten unseres Denkens oder unserer Anschauung befasst. Wir gehen naiv davon aus, dass man von je 2 Objekten x und y stets entscheiden kann, ob sie verschieden sind, oder ob x und y nur zwei verschiedene Namen desselben Objekts sind. In letzterem Fall schreiben wir $x = y$, sonst $x \neq y$. Eine **Menge** M ist nun eine Zusammenfassung verschiedener Objekte zu *einem* neuen Objekt. Ist x bei der Zusammenfassung von M mit aufgenommen worden, so nennen wir x ein **Element** von M und schreiben $x \in M$. Die Menge wird dadurch bestimmt welche Objekte zu ihr gehören, d.h. zwei Mengen M und N sind gleich, wenn sie dieselben Elemente enthalten. Formal:

$$M = N \quad \Longleftrightarrow \quad \Big(\text{für alle x gilt: } x \in M \iff x \in N\Big)$$

Wir können uns eine Menge also als Beutel vorstellen. Die Elemente sind die Objekte, die in dem Beutel sind. Dieses Bild hat nur 2 Fehler: (1) den Beutel selbst gibt es gar nicht, es geht nur um das 'wir gehören zu M' was die Elemente verbindet und (2) ein Objekt kann gleichzeitig zu vielen Mengen gehören, also in vielen Beuteln liegen.

Definiton: Es gibt 3 Schreibweisen, wie man eine Menge angeben kann: (1) direkt durch Aufzählung der Elemente. D.h. die Menge $\{ x_1, x_2, \dots, x_n \}$ besteht genau aus den Objekten x_1, x_2 bis x_n. Mehrfachnennungen sind möglich, die Reihenfolge ist unerheblich. (2) durch Angabe einer definierenden Eigenschaft. D.h. die Menge $\{ x \mid \varphi(x) \}$ besteht aus allen Objekten x, die die Eigenschaft $\varphi(x)$ erfüllen. Und (3) durch Auswahl aus einer bestehenden Menge M. D.h. die Menge $\{ x \in M \mid \varphi(x) \}$ besteht aus allen Elementen x von M, die zusätzlich die Eigenschaft $\varphi(x)$ erfüllen. Sind $n \in \mathbb{N}$ und $a < b \in \mathbb{R}$ so schreiben wir zum Beispiel

$$\begin{aligned} 1 \dots n &:= \{ k \in \mathbb{N} \mid 1 \leq k \text{ und } k \leq n \} \\ [a, b[&:= \{ x \in \mathbb{R} \mid a \leq x \text{ und } x < b \} \\ \emptyset &:= \{ x \mid x \neq x \} \end{aligned}$$

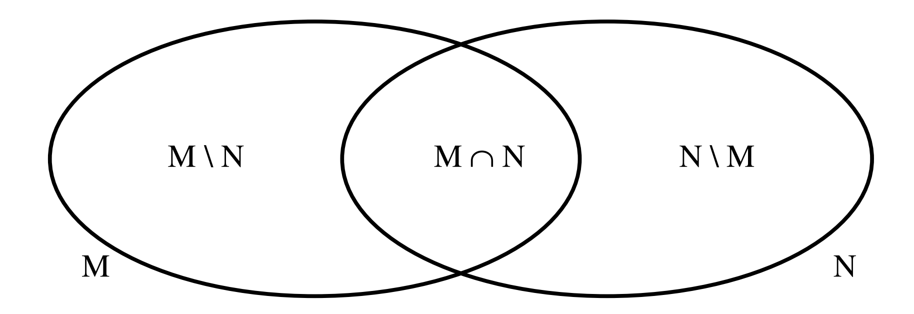
Die letzte Menge $\emptyset$ enthält nicht ein einziges Element, sie wird deshalb auch als **leere Menge** bezeichnet. Ist M eine beliebige Menge, dann schreiben wir $\#M$ für die Zahl der Elemente von M - man beachte, dass $\#M$ also eine Zahl (aus $\mathbb{N}$) oder unendlich ist. So ist zum Beispiel $\#(1\dots n) = n$ und $\#M = 0$ besagt nichts anderes, als $M = \emptyset$.

Mengenalgebra:

Definiton: Sind nun M und N zwei Mengen, dann definieren wir die **Vereinigung** $M \cup N$ als die Menge der x, die mindestens zu einem (M oder N) gehören. Der **Schnitt** $M \cap N$ ist analog die Menge der x, die zu beiden (M und N) gehören und die **Differenz** $N \setminus M$ besteht aus den Objekten von N, die nicht (auch) zu M gehören. Formal:

$$\begin{aligned} M \cup N &:= \{\, x \mid x \in M \text{ oder } x \in N \,\} \\ M \cap N &:= \{\, x \mid x \in M \text{ und } x \in N \,\} \\ N \setminus M &:= \{\, x \mid x \in N \text{ und } x \notin M \,\} \end{aligned}$$

Eine bekannte Darstellung sind die Venn-Diagramme, bei denen jede Menge durch einen (eiförmigen) Kreis dargestellt wird. Die Vereinigung $M \cup N$ ist dann die Gesamtfläche, der Schnitt $M \cap N$ ist der Überlapp und die Differenz $N \setminus M$ ist N ohne den Überlapp mit M:



Venn-Diagramm der Vereinigung $M \cup N$

Anhand dieser Definitionen sieht man dann leicht ein, dass die folgenden *de Morgan'schen Regeln* gelten (zeichnen Sie die jeweiligen Venn-Diagramme):

$$\begin{aligned} M &= M \cap (M \cup N) \\ M &= M \cup (M \cap N) \\ L \cup (M \cap N) &= (L \cup M) \cap (L \cup N) \\ L \cap (M \cup N) &= (L \cap M) \cup (L \cap N) \end{aligned}$$

Es kommt häufig vor, dass die vorliegenden Daten strukturiert sind. Man kann zum Beispiel die Körpergröße 183 (cm) und das Körpergewicht 78 (kg) betrachten. Will man dem Rechnung tragen, dass *eine* Person diese *beiden* Maße besitzt, so gruppiert man Größe und Gewicht zu einem Datensatz $(183, 78)$. Zu jeder Person gehört also Ihr eigener Datensatz. Wir führen diese Konstruktion nun allgemein aus: sind x_1, x_2, bis x_n irgendwelche Objekte, dann können wir diese zu einer Liste $x = (x_1, x_2, \ldots, x_n)$ (genannt n-**Tupel**) zusammen fassen. Dieses Tupel x ist dann *ein* neues Objekt. Zwei n-Tupel $x = (x_1, x_2, \ldots, x_n)$ und $y = (y_1, y_2, \ldots, y_n)$ sind genau dann gleich, wenn Sie in allen Eintrgën überein stimmen, formal:

$$x = y \quad \Longleftrightarrow \quad x_1 = y_1 \text{ und } x_2 = y_2 \text{ und } \ldots \text{ und } x_n = y_n$$

Definiton: Sind nun M_1 und $M_2, \ldots$ bis M_n irgendwelche Mengen, dann können wir alle Kombinationen $(x_1, x_2, \ldots, x_n)$ von n-Tupeln bilden, wobei x_1 aus M_1 stammt, x_2 aus M_2 und so weiter, bis x_n aus M_n. Die Menge, die aus all diesen Kombinationen besteht, heißt **Karthesisches Produkt** der Mengen M_1, M_2 bis und M_n und wir schreiben sie als:

$$M_1 \times M_2 \times \cdots \times M_n \quad := \quad \{\, (x_1, x_2, \ldots, x_n) \mid x_i \in M_i \ (i \in 1 \ldots n) \,\}$$

Beispiel: Als Beispiel betrachten wir die Mengen $M = \{\, a, b \,\}$ und $N = \{\, 1, 2, 3 \,\}$. Das Karthesische Produkt dieser Mengen lautet nun explizit:

$$\{\, (a, 1),\ (a, 2),\ (a, 3),\ (b, 1),\ (b, 2),\ (b, 3) \,\}$$

An diesem Beispiel wird klar: enthält M_1 genau a_1 Elemente, enthält M_2 genau a_2 Elemente und so fort, enthält M_n genau a_n Elemente, dann enthält das karthesische Produkt genau $a_1 a_2 \ldots a_n$ Elemente. Formal:

$$\#(M_1 \times M_2 \times \cdots \times M_n) \quad = \quad \prod_{i=1}^{n} \#M_i$$

Teilmengen:

Definiton: Wir nennen M eine **Teilmenge** von N, falls M in N enthalten ist. Genauer, wenn jedes Element x aus M auch in N liegt. Formal also:

$$M \subseteq N \quad :\Longleftrightarrow \quad \Big(\text{für alle x gilt: } x \in M \implies x \in N\Big)$$

Offenbar ist die Teilmengenrelation eine Halbordnung. D.h. ist L eine Teilmenge von M und M eine Teilmenge von N, dann ist L erst recht eine Teilmenge von N. Und die Mengen M und N sind genau dann gleich $M = N$, wenn $M \subseteq N$ und $N \subseteq M$ sich gegenseitig enthalten. Formal geschrieben:

$$\begin{aligned} L \subseteq M \text{ und } M \subseteq N \quad &\Longrightarrow \quad L \subseteq N \\ M \subseteq N \text{ und } N \subseteq M \quad &\Longleftrightarrow \quad M = N \end{aligned}$$

Das einfachste Beispiel der Teilmengeneigenschaft ist $M \cap N \subseteq M \subseteq M \cup N$ für beliebige Mengen M und N. Genauso leicht sieht man die Äquivalenzen

$$M \subseteq N \iff M \setminus N = \emptyset \iff M = M \cap N \iff N = M \cup N$$

Ein weiteres Beispiel für Teilmengen sind die bekannten Zahlenmengen. Diese sind hintereinander geordnet: $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. Wir nutzen die Gelegenheit um an die Bedeutung der Zahlenmengen zu erinnern:

Name	Symbol	Beispiele	allgemein
Natürliche Zahlen	$\mathbb{N}$	$0, 4, 17, \ldots$	$0, 1, 2, 3, \ldots$
Ganze Zahlen	$\mathbb{Z}$	$-4, 0, 17, \ldots$	$a - b$, mit $a, b \in \mathbb{N}$
Rationale Zahlen	$\mathbb{Q}$	$\frac{2}{7}, 3, \frac{28}{6}, \ldots$	$\frac{a}{b}$, mit $a, b \in \mathbb{Z}$, $b \neq 0$
Reelle Zahlen	$\mathbb{R}$	$\sqrt{2}, \pi, e, \ldots$	Dezimalbrüche
Komplexe Zahlen	$\mathbb{C}$	$1 + i, \pi - 2i, \ldots$	$a + ib$ mit $a, b \in \mathbb{R}$

Definiton: Es kommt gelegentlich vor, dass man alle Teilmengen einer Menge M betrachten möchte. Dann braucht man die Menge aller Teilmengen von M und diese wird **Potenzmenge** von M genannt:

$$\mathcal{P}(M) \; := \; \{ A \mid A \subseteq M \}$$

Offensichtlich ist jede Menge in sich selbst enthalten $M \subseteq M$, also ist immer $M \in \mathcal{P}(M)$. Und da $\emptyset$ gar keine Elemente enthält, gilt auch $\emptyset \subseteq M$ also $\emptyset \in \mathcal{P}(M)$. Ist $M = \emptyset$ so sind das auch schon alle Teilmengen, und damit gilt $\mathcal{P}(\emptyset) = \{ \emptyset \}$. Allgemein gilt: besitzt die Menge M genau n Elemente, dann besitzt die Potenzmenge 2^n Elemente, formal: $\#\mathcal{P}(M) = 2^{\#M}$. Wir führen das einmal an dem Beispiel der Menge $M = \{ 2, 3, 5 \}$ aus:

$$\mathcal{P}(M) \; = \; \{ \emptyset, \{ 2 \}, \{ 3 \}, \{ 5 \}, \{ 2, 3 \}, \{ 2, 5 \}, \{ 3, 5 \}, M \}$$

Definiton: Hat man eine äussere Menge M *fixiert* und ist $A \subseteq M$ eine Teilmenge, so definert man das **Komplement** $\overline{A}$ von A als den Teil von M, der nicht schon in A enthalten ist. Formal also:

$$\overline{A} \; := \; M \setminus A$$

Sind nun A und $B \subseteq M$ beliebige Teilmengen von M, so sieht man (am einfachsten wieder mit Venn-Diagrammen) die folgenden Eigenschaften des Komplements ein (die ebenfalls zu den *de Morgan'schen Regeln* gehören):

$$\begin{aligned} A \cap \overline{A} &= \emptyset \\ A \cup \overline{A} &= M \\ \overline{A \cup B} &= \overline{A} \cap \overline{B} \\ \overline{A \cap B} &= \overline{A} \cup \overline{B} \end{aligned}$$

Kapitel 3

Summen und Produkte

Definiton: Sind x_1, x_2 bis $x_n \in \mathbb{R}$ beliebige (reelle) Zahlen, dann definiert man die folgenden Schreibweisen (genannt **Summe** bzw. **Produkt** der x_k):

$$\begin{aligned} \sum_{k=1}^{n} x_k &:= x_1 + x_2 + \cdots + x_k \\ \prod_{k=1}^{n} x_k &:= x_1 \cdot x_2 \cdot \ldots \cdot x_k \end{aligned}$$

Satz: Man bemerke, dass die rechte Seite des Ausdrucks ohne Klammern auskommt, da das Ergebnis immer gleich sein wird, egal wie die Klammern gesetzt würden (Assoziativität). Die üblichen Rechenregeln (Kommutativität und Distributivität) ergeben dann eine Liste von unmittelbar einsichtigen Rechenregeln für Summen (und analog für Produkte)

$$\begin{aligned} \sum_{k=1}^{n} a x_k &= a \sum_{k=1}^{n} x_k \\ \sum_{k=1}^{n} (x_k + y_k) &= \sum_{k=1}^{n} x_k + \sum_{k=1}^{n} y_k \\ \sum_{k=0}^{n-1} x_k &= \sum_{k=1}^{n} x_{k-1} \\ \sum_{k=1}^{n+1} x_k &= \sum_{k=1}^{n} x_k + x_{n+1} \\ \left(\sum_{i=1}^{m} x_i \right) \left(\sum_{j=1}^{n} y_j \right) &= \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j \end{aligned}$$

Und die bekannten Rechenregeln für die Exponential- und Logarithmusfunktion drücken sich damit wie folgt aus:

$$\begin{aligned}\exp\left(\sum_{k=1}^{n} x_k\right) &= \prod_{k=1}^{n} \exp(x_k) \\ \ln\left(\prod_{k=1}^{n} x_k\right) &= \sum_{k=1}^{n} \ln(x_k)\end{aligned}$$

Oft ist x_k eine einfache Funktion von k, z.B. $x_k = k$ oder $x_k = k^2$. In diesen Fällen ist die Summenschreibweise besonders hilfreich. Wir möchten ein einfaches Beispiel für diese Situation (in einer Doppelsumme) anfügen:

$$\begin{aligned}\sum_{i=1}^{3}\sum_{j=1}^{4}(i+j) &= \sum_{j=1}^{4}(1+j) + \sum_{j=1}^{4}(2+j)\sum_{j=1}^{4}(3+j) \\ &= (2+3+4+5)+(3+4+5+6)+(4+5+6+7) = 54\end{aligned}$$

Satz: Schließlich geben wir noch ein paar Formeln für spezielle Summen an: unter anderem ergibt die Summe der ersten n Zahlen $(n+1)n/2$, die Summe der ersten n ungeraden Zahlen ergibt n^2 und die Summe der ersten n Quadratzahlen ergibt $(2n+1)(n+1)n/6$:

$$\begin{aligned}\sum_{k=1}^{n} a &= na \\ \sum_{k=1}^{n} k &= \frac{n(n+1)}{2} \\ \sum_{k=1}^{n}(2k-1) &= n^2 \\ \sum_{k=0}^{n} q^k &= \frac{q^{n+1}-1}{q-1} \\ \sum_{k=1}^{n} k^2 &= \frac{(2n+1)(n+1)n}{6}\end{aligned}$$

Kapitel 4

Prozentrechnung

Wenn wir von einem Wachstum um 100% sprechen, meinen wir dass sich der betrachtete Betrag verdoppelt hat. Und legen wir 1000 Euro bei einem Zinssatz von 10% (pro Jahr) an, so haben wir nach einem Jahr einen Kontostand von 1100 Euro. Ein Prozentwert soll also ein Verhältnis zwischen dem Anfangs- und dem Endwert ausdrücken. Und die Rechenvorschrift sieht dabei folgendermaßen aus:

$$\begin{aligned}\text{Wachstumsfaktor} &= 1 + \frac{\text{Zinssatz}}{100} \\ \text{Endwert} &= \text{Wachstumsfaktor} \cdot \text{Anfangswert}\end{aligned}$$

Etwas formaler: im folgenden bezeichnen wir den Anfangswert mit x, den Endwert mit x'. Bezeichnen wir weiterhin den Zinssatz mit z, dann ist der Wachstumsfaktor gegeben, durch $q = 1 + z/100$. Und der Endwert ist dann $x' = qx$. Irreführender Weise wird ein prozentuales Wachstum in der wirtschaftlichen Praxis aber additiv geschrieben, man meint dabei

$$x + z\% \quad := \quad \left(1 + \frac{z}{100}\right) x$$

Daran sieht man: ob ein Betrag erst um $w\%$ und danach um $z\%$ wächst, oder umgekehrt, erst um $z\%$ und danach um $w\%$, macht keinen Unterschied (denn die Multiplikation ist kommutativ). Formal heißt das

$$(x + w\%) + z\% \quad = \quad (x + z\%) + w\%$$

Aber Vorsicht: legen wir die 1000 erst bei 5% und danach mit 10% an, so ernten wir am Ende den Betrag $1,05 \cdot 1,1 \cdot 1000 = 1150$. Das entspricht einem Zinssatz von $15,5\%$ und eben nicht $5 + 10 = 15$ Prozent. Es ist also

$$(x + w\%) + z\% \quad \neq \quad x + (w + z)\%$$

Wie bestimmt man eigentlich den Zinssatz? Löst man die Gleichungen $x' = qx$ nach q und $q = 1 + z/100$ nach z auf, so findet man $q = x'/x$ und $z = 100(q-1)$, also

$$\begin{aligned} \text{Wachstumsfaktor} &= \frac{\text{Endwert}}{\text{Anfangswert}} \\ \text{Zinssatz} &= 100 \cdot (\text{Wachstumsfaktor} - 1) \end{aligned}$$

Problem 1: Der Anfangsbetrag $x \in \mathbb{R}$ wird mit dem Zinssatz $z \in \mathbb{R}$ verzinst. Wie groß ist der Endbetrag x' nach $n \in \mathbb{N}$ Zinsperioden? Klar:

$$x' = \left(1 + \frac{z}{100}\right)^n x$$

Problem 2: Wir verzinsen 1000 Euro 2 Jahre lang mit 10%. Welchen monatlichen Zinssatz müsste man nehmen, um denselben Endbetrag (1210 Euro) zu erreichen? Die Lösung finden wir aus der Gleichung $1000 \cdot (1,1)^2 = 1000 \cdot q^{24}$ zu $q = 1,0079\ldots$ also etwa $z = 0,8\%$. Allgemeiner lautet das Problem: wir verzinsen x über m Perioden mit dem Zinssatz w. Welchen Zinssatz z muss man wählen, um denselben Endbetrag in n Perioden zu erreichen? Die Lösung erhält man wieder durch Auflösen der Gleichung $x(1 + w/100)^m = x(1 + z/100)^n$ nach z zu

$$z = 100 \cdot \left(\left(1 + \frac{w}{100}\right)^{\frac{m}{n}} - 1\right)$$

Problem 3: Zu Beginn jeder Periode wird der Betrag $x \in \mathbb{R}$ auf ein Konto eingezahlt und (am Ende der Periode) wird alles mit dem Zinssatz $z \in \mathbb{R}$ verzinst. Wie hoch ist der Endbetrag nach $n \in \mathbb{N}$ Perioden? Nach einer Periode ist x einmal verzinst worden und x wurde ein weiteres Mal eingezahlt, also $x' = qx + x = (q+1)x$. Nach der zweiten Periode wurde dieser Betrag wieder verzinst und ein weiteres Mal wurde x eingezahlt, also $x' = q(qx + x) + x = (q^2 + q + 1)x$. So geht das immer fort, bis nach n Perioden $x' = (q^n + q^{n-1} + \cdots + q + 1)x$ ist. Nach den Summenformeln aus Kapitel 3 also

$$x' = \frac{q^{n+1} - 1}{q - 1} x \quad \text{wobei} \quad q = 1 + \frac{z}{100}$$

Kontinuierliche Verzinsung:

Beispiel: Wir legen 1000 Euro bei einem jählichen Zinssatz von 4% an. Nach 3 Jahren und 3 Monaten (also $3,25$ Jahren) wollen wir unser Konto wieder auflösen. Wieviel Geld liegt jetzt auf dem Konto? In der Finanzwirtschaft geht man nun folgender Maßen vor: zunächst berechnet man die Verzinsung für die 3 ganzen Jahre, also $(1,04)^3 \cdot 1000 = 1124,86$ Euro. Dieser Betrag lag ja noch ein viertel Jahr bei 4% pro Jahr auf dem Konto. Dann wird der Einfachheit halber einfach noch ein Viertel der jährlichen Verzinsung genommen, also $x' = (1 + 0,25 \cdot 0,04) \cdot 1124,86 = 1113,73$. Bei der sogenannten *kaufmännischen Verzinsung* wird die Zeitspanne t also aufgeteilt in $t = n+r$, wobei $n \in \mathbb{Z}$ ganzzahlig und r der Rest $0 \le r < 1$ ist. Es bezeichne wieder $q = 1 + z/100$ und $p = q - 1 = z/100$. Dann ist der Endbetrag, also

$$x' \;=\; q^n(1+rp)x \;=\; \left(1+\frac{z}{100}\right)^n \left(1+\frac{rz}{100}\right) x$$

Aber eigentlich macht man da einen Fehler: wegen der Zinseszins-Effekte wächst der Betrag am Ende ja schneller als am Anfang. Die *Änderung* des Betrages ist ja proportional zum bisher bestehenden Betrag (je mehr Geld auf dem Konto ist, desto mehr wird auch verzinst). Legt man 1000 Euro ein Jahr lang bei 100% an erhält man bei jählicher Verzinsung 2000 Euro zurück. Hätte man stattdessen das Geld jede Woche abgehoben und gleich wieder angelegt, hätte man (bei kaufmännischer Verzinsung) einen Zinssatz von $(100/52)\% = 1,92\%$ erhalten und aus den 1000 Euro wären satte $1,0192^{52} \cdot 1000 = 26130$ Euro geworden. Macht man die Dauer einer Zinsperiode immer kürzer nähert sich dieser Betrag immer weiter $1000e = 27182,81\ldots$ Euro. Es ist unbefriedigend, dass der Endbetrag gesteigert werden kann, wenn man die Zinsperiode verkürzt.

Wir wollen diesen Makel der kaufmännischen Verzinsung also bereinigen. Die Idee ist die Dauer einer Zinsperiode schon im Vorfeld rechnerisch immer weiter zu verkleinern. Also statt 100% jährlich, lieber 50% halbjährlich, oder noch besser 25% vierteljährlich und so weiter. Betrachten wir was dabei heraus kommt, wenn man die Schritte immer weiter verfeinert:

$$\begin{aligned} e(x) &:= \lim_{n\to\infty} \left(1+\frac{x}{n}\right)^n \\ &= 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \cdots + \frac{1}{1\cdot 2\cdot 3\cdot \ldots \cdot n}x^n + \cdots \end{aligned}$$

Diese Funktion nennt man *Exponentialfunktion*. Anhand der Darstellung als unendlich lange Summe sieht man, dass $e'(x) = e(x)$ ist. Bei Ihr ist die Änderung also immer gleich dem gegenwärtigen Betrag. Die Zahl $e := e(1) = 2,7182\ldots$ ist eine der wichtigen mathematischen Konstanten, ähnlich wie $\pi = 3,1415\ldots$. Durch die Eigenschaft eine Summe in ein Produkt zu verwandeln $e(x+y) = e(x)e(y)$ ergibt sich, dass $e(k) = e^k$ (für $k \in \mathbb{Z}$) ist. Deswegen schreibt man auch $e^x := e(x)$ für $x \in \mathbb{R}$. Mit Hilfe dieser Funktion lässt sich die kaufmännische Verzinsung dann verbessern, zur *kontinuierlichen Verzinsung*:

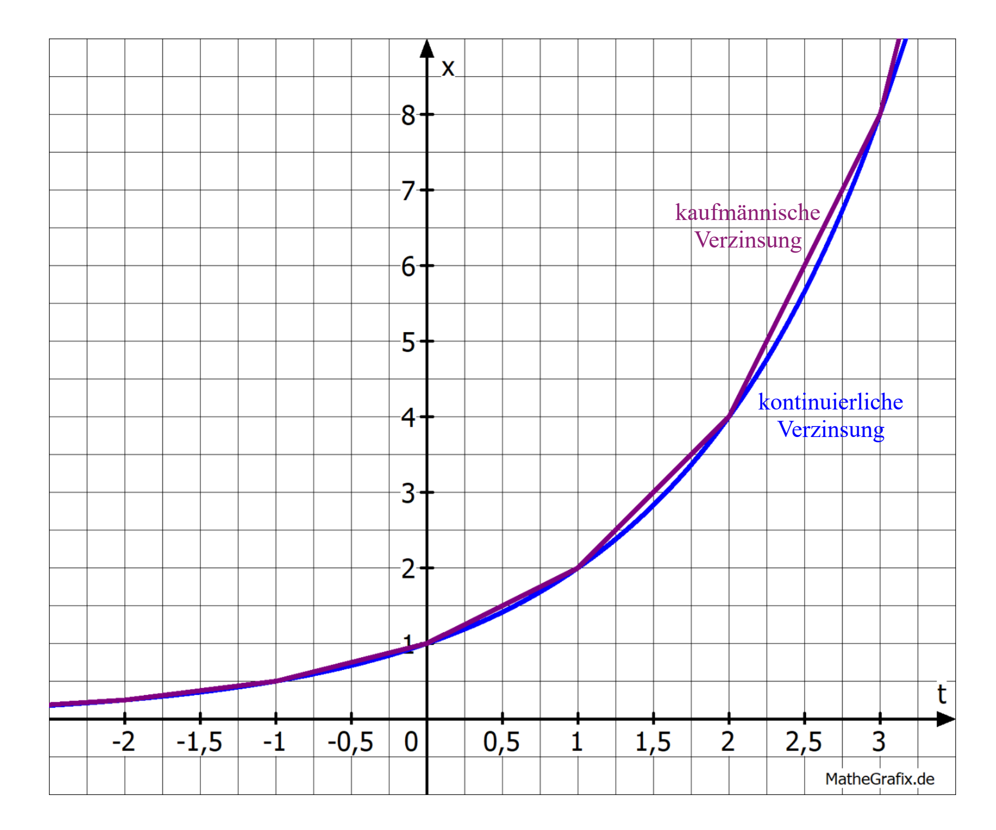
$$x' \;=\; q^t x \;:=\; e^{t\cdot\ln(q)}$$

Dabei ist *ln* der *(natürliche) Logarithmus*, die Umkehrfunktion der e-Funktion. Diese kann man mit (für $-1 < x \leq 1$) Hilfe einer anderen, unendlichen Summe berechnen

$$\ln(1+x) \quad = \quad x - \frac{x^2}{2} + \frac{x^3}{3} + \cdots + (-1)^{n-1}\frac{x^n}{n} + \cdots$$

Setzt man diese beiden Reihen in kontinuierliche Verzinsung $x' = q^t x$ ein, sieht man, dass die kaufmännische Verzinsung einfach die lineare Approximation der kontinuierlichen Verzinsung ist:

$$q^t x \quad = \quad q^n q^r x \quad = \quad q^n\Big(1 + r\Big(p + \cdots\Big) + \cdots\Big)x \quad \approx \quad q^n(1+rp)x$$



Problem 4: Der Betrag $x \in \mathbb{R}$ wird zum Zinssatz $z \in \mathbb{R}$ angelegt. Wie lange muss x (kontinuerlich) verzinst werden, um den Betrag $x' \in \mathbb{R}$ zu erreichen? Dazu müssen wir die Gleichung $x' = q^t x$ nach t auflösen: $\ln(x') = \ln(q^t x) = t \cdot \ln(q) + \ln(x)$, also

$$t \quad = \quad \frac{\ln(x') - \ln(x)}{\ln(q)}$$

Bemerkung: Dasselbe Problem ist im Falle der kaufmännischen Verzinsung leider nicht exakt lösbar. Sei wieder $p = z/100$ und wir suchen den Zinssatz z, dann muss man die folgende Gleichung (numerisch, z.B. durch Intervallhalbierung) nach p lösen

$$(1+p)^n(1+rp) \quad = \quad \frac{x'}{x}$$

Kapitel 5

Vektoralgebra

Wie üblich bezeichnet $\mathbb{R}$ die Menge der reellen Zahlen [rein algebraisch gesehen, könnte die folgende Konstruktion mit einem beliebigen kommutativen Ring R ausgeführt werden]. Die Menge $\mathbb{R}^n$ besteht dann aus Listen $x = (x_1, x_2, \dots, x_n)$ von n reellen Zahlen x_1, x_2 bis $x_n \in \mathbb{R}$. Eine solche Liste heißt auch **Vektor**. Wir verabreden, dass die Liste $x \in \mathbb{R}^n$ immer die Einträge (**Komponenten**) $x_t \in \mathbb{R}$ hat. Analog besteht $y \in \mathbb{R}^n$ immer aus den Komponenten y_t, d.h. $y = (y_1, y_2, \dots, y_n)$. Zwei Vektoren x, $y \in \mathbb{R}^n$ sind gleich, wenn sie in allen Komponenten überein stimmen, formal:

$$x = y \quad \Longleftrightarrow \quad x_1 = y_1 \text{ und } x_2 = y_2 \text{ und } \dots \text{ und } x_n = y_n$$

Wir interpretieren eine Zahl $a \in \mathbb{R}$ immer auch gleich als Vektor, indem wir einfach jede Komponente des Vektors als a nehmen. D.h. wir identifizieren

$$a \in \mathbb{R} \quad = \quad (a, a, \dots, a) \in \mathbb{R}^n$$

Definiton: Nun definieren wir drei verschiedene Rechenoperationen auf $\mathbb{R}^n$: seien also wieder x, $y \in \mathbb{R}^n$ zwei Vektoren. Dann erklären wir die **Vektoraddition** $x + y \in \mathbb{R}^n$, die **Vektormultiplikation** $xy \in \mathbb{R}^n$ und das **Skalarprodukt** $\langle x \mid y \rangle \in \mathbb{R}$ durch

$$\begin{aligned} x + y &:= (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n) \\ xy &:= (x_1 y_1,\ x_2 y_2,\ \dots,\ x_n y_n) \\ \langle x \mid y \rangle &:= x_1 y_1 + x_2 y_2 + \dots + x_n y_n \end{aligned}$$

Man beachte, dass aufgrund der Einbettung $a = (a, a, \dots, a)$ und der Vektormultiplikation xy auch gleich die **Skalarmultiplikation** ax definiert wurde:

$$ax \quad = \quad (ax_1, ax_2, \dots, ax_n)$$

Vom abstrakten Standpunkt gesehen wird $\mathbb{R}^n$ damit zu einer $\mathbb{R}$-Algebra, also insbesondere zu einem $\mathbb{R}$-Vektorraum. Anschaulich, geometrisch ist ein Vektor $x = (x_1, x_2, \ldots, x_n)$ aber einfach ein Punkt im n-dimensionalen Raum. Zum Beispiel ist (x_1, x_2) ein Punkt in der Ebene und (x_1, x_2, x_2) ein Punkt im Raum. Der Abstand der Punkte x und y ist nach Pythagoras $d_2(x, y)$ (s. unten). Es gibt aber noch weitere Definitionen, die einen sinnvollen Abstandsbegriff liefern, unter anderem:

$$\begin{aligned} d_1(x,y) &:= \sum_{t=1}^{n} |x_t - y_t| \\ d_2(x,y) &:= \sqrt{\sum_{t=1}^{n} (x_t - y_t)^2} \\ d_\infty(x,y) &:= \max\{\, |x_1 - y_1|, \ldots, |x_n - y_n| \,\} \end{aligned}$$

Natürlich liefern die verschiedenen Abstandsbegriffe auch verschiedene Zahlenwerte. Es gibt aber ein paar wichtige Abschätzungen, die allgemein gelten:

$$\frac{1}{\sqrt{n}}\, d_2(x,y) \;\leq\; d_\infty(x,y) \;\leq\; d_1(x,y) \;\leq\; \sqrt{n}\, d_2(x,y)$$

Wir werden im nächsten Kapitel auch den **Durchschnitt** $A(x)$ der Zahlen x_1, x_2 bis x_n einführen. Dieser wird definiert werden, als:

$$A(x) := \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

Satz: Man beachte, dass der Durchschnitt eines Vektorprodukts xy und das Skalarprodukt zweier Vektoren x und y eng verwandt sind. So werden wir die folgenden Identitäten häufig benutzen:

$$\begin{aligned} A(xy) &= \frac{1}{n}\sum_{t=1}^{n} x_t y_t &= \frac{1}{n}\langle x \mid y \rangle \\ A(x^2) &= \frac{1}{n}\sum_{t=1}^{n} x_t^2 &= \frac{1}{n}\langle x \mid x \rangle \end{aligned}$$

Besonders hilfreich ist diese Notation, wenn wir den Vektor $t = (1, 2, \ldots, n)$ einführen. Dies ist zwar formal nicht ganz korrekt, da t ja auch der Index (von 1 bis n) sein soll, führt aber zu der suggestiven Schreibweise:

$$A(ty) = \frac{1}{n}\sum_{t=1}^{n} t y_t$$

Kapitel 6

Eindimensionale Verteilungen

Wir betrachten n Merkmalsträger, durchnummeriert mit den Zahlen $1, 2, \ldots$ bis n. Da wir nur ein einziges Merkmal x betrachten, gehört zu jedem Merkmalsträger t nur eine *Ausprägung* x_t. Die Menge aller möglichen Ausprägungen bezeichnen wir mit S. Wir haben es bei dem Merkmal x also mit einer Zuordnung zu tun, der Form:

$$x \, : \, 1 \ldots n \to S \, : \, t \mapsto x_t$$

Zumeist gibt es nur endlich viele verschiedene Ausprägungen $s_1, s_2, \ldots, s_m$ des Merkmals x, dann ist also $S = \{ s_1, s_2, \ldots, s_m \}$. Gelegentlich kommt aber auch $S = \mathbb{N}$ (diskret, positiv, z.B. Umsätze in Cent), $S = \mathbb{Z}$ (diskret, z.B. Kontostände in Cent), $S = \mathbb{R}^+$ (kontinuierlich, positiv, z.B. Wechselkurse) oder sogar $S = \mathbb{R}$ (kontinuierlich, z.B. Zeitunterschiede) vor. Zu $s = s_i$ bezeichnen wir nun die *absolute Häufigkeit*, mit der s_i unter den x_t vorkommt mit n_i. Die *relative Häufigkeit* bezeichnen wir mit h_i, formal lautet dies $(i \in 1 \ldots m)$

$$\begin{aligned} n_i &:= \# \{ t \in 1 \ldots n \mid x_t = s_i \} \\ h_i &:= \frac{n_i}{n} \end{aligned}$$

Kann man die s_i vergleichen (komparatives Merkmal, z.B. Handelsklassen) geht man immer von einer aufsteigenden Sortierung $s_1 \leq s_2 \leq \cdots \leq s_m$ der möglichen Ausprgüngen aus. Dann kann man auch die *kumulierten, absoluten* bzw. *relativen Häufigkeiten* einführen $(k \in 1 \ldots m)$:

$$\begin{aligned} N_k &:= \sum_{i=1}^{k} n_i = N_{k-1} + n_k \\ H_k &:= \frac{N_k}{n} = \sum_{i=1}^{k} h_i = H_{k-1} + h_k \end{aligned}$$

Der Vollständigkeit halber setzt man daher auch $N_0 := 0$ und $H_0 := 0$. Ist $S \subseteq \mathbb{R}$ (quantitatives Merkmal, z.B. Kontostände), kann man die Ausprägungen auch selbst addieren, um die Größe des Bestandes auszuwerten, man bezeichnet ($i \in 1 \dots m$):

$$\begin{aligned} X_i &:= n_i s_i \\ X &:= \sum_{t=1}^{n} x_t = \sum_{i=1}^{m} n_i s_i = \sum_{i=1}^{m} X_i \end{aligned}$$

Dabei heißt X_i die *absolute Merkmalssumme* von s_i, während der Gesamtbetrag X auch *totale Merkmalssumme* heißt. Der Durchschnitt (genauer gesagt, das *arithmetische Mittel*) ist bekanntlich der Gesamtbetrag pro Person, also

$$A(x) := \frac{1}{n} X$$

Für die Konzentrationsanalyse betrachtet man schließlich die noch Anteile am Gesamtbetrag, man nennt dies die *relative Merkmalssumme* ℓ_i bzw. die *kumulierte, relative Merkmalssumme* L_k ($i, k \in 1 \dots m$, wieder setzt man der Vollständigkeit halber $L_0 := 0$):

$$\begin{aligned} \ell_i &:= \frac{X_i}{X} = \frac{h_i s_i}{A(x)} \\ L_k &:= \sum_{i=1}^{k} \ell_i = L_{k-1} + \ell_k \end{aligned}$$

Lagemaße:

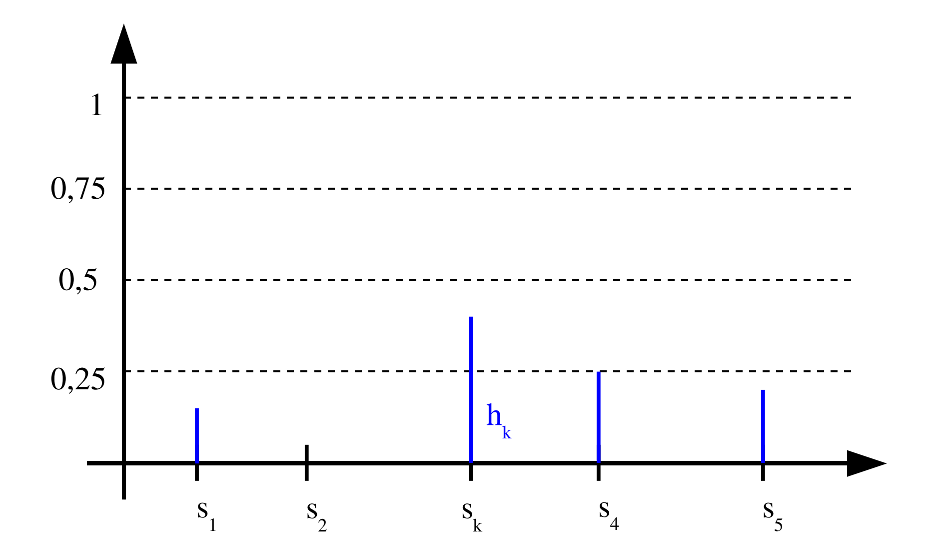
Die Statistik ist daran interessiert eine Flut von Daten $x_1, x_2, \dots, x_n$ zu komprimieren. Und als erstes will man so etwas wie den typischen Vertreter der x_t haben. Dieser typische Vertreter soll die Größe der x_t wiedergeben. Da er also sagt, *wo* die Daten liegen, spricht man auch von einem *Lagemaß*. Die einfachste Möglichkeit ist es, den Wert zu nehmen, der unter den x_t am häufigsten vorkommt. Diesen nennt man den **Modus** (oder auch *dichtester Wert*) $D(x)$:

$$D(x) := s_k \quad \text{so dass} \quad n_k = \max\{\, n_i \mid i \in 1 \dots m \,\}$$

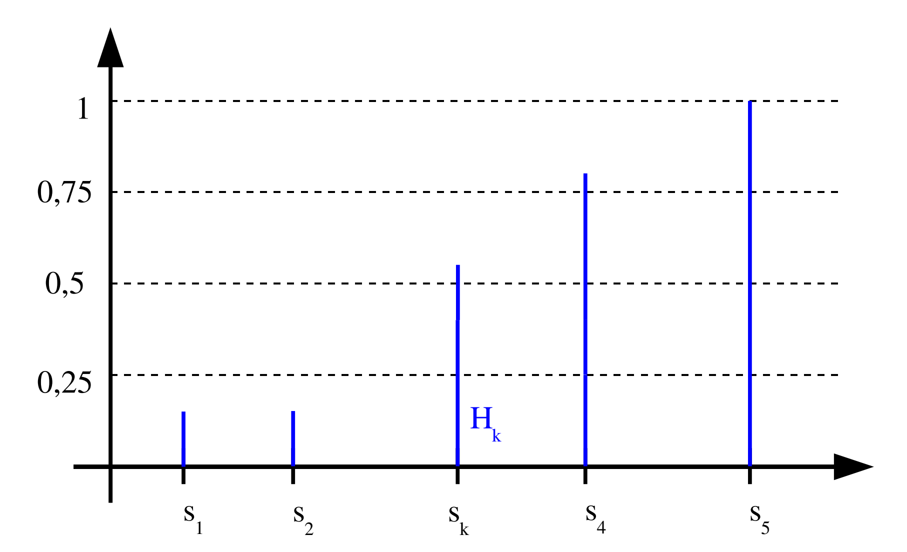
Bei einem komparativen Merkmal x sortiert man die x_t aufsteigend der Größe nach: $x_1 \leq x_2 \leq \cdots \leq x_n$. Der Wert, der dann in der Mitte steht, heißt **Median** (oder auch *Zentralwert*) $Z(x)$:

$$Z(x) := x_k \quad \text{wobei} \quad k = \begin{cases} \frac{n}{2} & \text{für } n \text{ gerade} \\ \frac{n+1}{2} & \text{für } n \text{ ungerade} \end{cases}$$

In einem **Balkendiagramm** stellt man die Häufigkeiten graphisch dar, indem man die Balken unter die Punkte $(s_i \mid h_i)$ einzeichnet. Der Modus ist also dasjenige s_k mit dem höchsten Balken. Der Median hingegen ist der Wert s_k, an dem die Stufenfunktion zu den Punkten $(s_i \mid H_i)$ erstmals den Wert $1/2$ erreicht bzw. überschreitet.



Balkendiagramm mit $m = 5$, $k = 3$,
$h_1 = 0.15$, $h_2 = 0$, $h_3 = 0.4$, $h_4 = 0.25$ und $h_5 = 0.2$



kumuliertes Balkendiagramm für dieselben Werten

So wie der Median derjenige Wert ist, der den Übergang von der unteren zur oberen Hälfte markiert, kann man natürlich andere Trennlinien betrachten: das **erste Quartil** $Q_1(x)$ markiert die Grenze zwischen dem unteren Viertel und den oberen drei Vierteln. Und das **dritte Quartil** $Q_3(x)$ markiert umgekehrt die Grenze zwischenden unteren drei Vierteln und dem oberen Viertel. Das *zweite Quartil* ist genau der Median $Q_2(x) = Z(x)$. Formal sind die Quartile also folgenderma/ssen zu definieren:

$$Q_j(x) \quad := \quad s_k \quad \text{wobei} \quad H_{k-1} < \frac{j}{4} \leq H_k$$

Das bekannteste Lagemaß ist sicherlich das **arithmetische Mittel** (oder auch *Durchschnitt*) $A(x)$. Es ist die Gesamtsumme X pro Person, kann also nur bei quantitativen Merkmalen x gebildet werden:

$$A(x) \quad := \quad \frac{1}{n}\sum_{t=1}^{n} x_t \quad = \quad \frac{1}{n}X \quad = \quad \sum_{i=1}^{m} h_i s_i$$

Anhand dieser Definition rechnet man leicht nach, dass das arithmetische Mittel eine Reihe nützlicher Eigenschaften erfüllt ($a \in \mathbb{R}$, x und $y \in \mathbb{R}^n$):

$$\begin{aligned} A(a) &= a \\ A(ax) &= aA(x) \\ A(x+y) &= A(x) + A(y) \end{aligned}$$

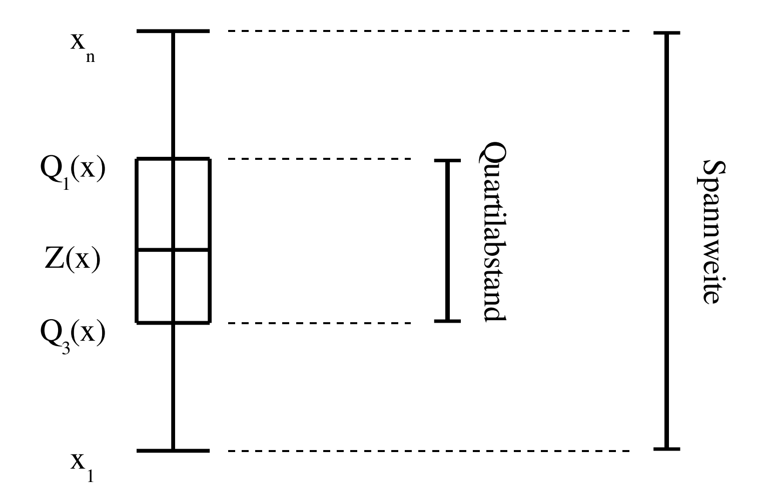
Bemerkung: Der Zentralwert und der Durchschnitt haben eine strukturelle Gemeinsamkeit: beide entstehen durch Projektion des Datenpunktes x auf die Diagonale. Genauer: wir wollen die n Zahlen x_1 bis x_n durch eine Zahl α ersetzen. Fasst man $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ als Punkt im n-dimensionalen Raum auf, so sucht man das α, so dass der Punkt $\alpha = (\alpha, \alpha, \ldots, \alpha) \in \mathbb{R}^n$ möglichst dicht bei x liegt. Nun kann man den Abstand zwischen 2 Punkten x und y im $\mathbb{R}^n$ aber auf verschiedene Weisen messen, unter anderem mit:

$$\begin{aligned} d_1(x,y) &= \sum_{t=1}^{n} |x_t - y_t| \\ d_2(x,y) &= \sqrt{\sum_{t=1}^{n} (x_t - y_t)^2} \end{aligned}$$

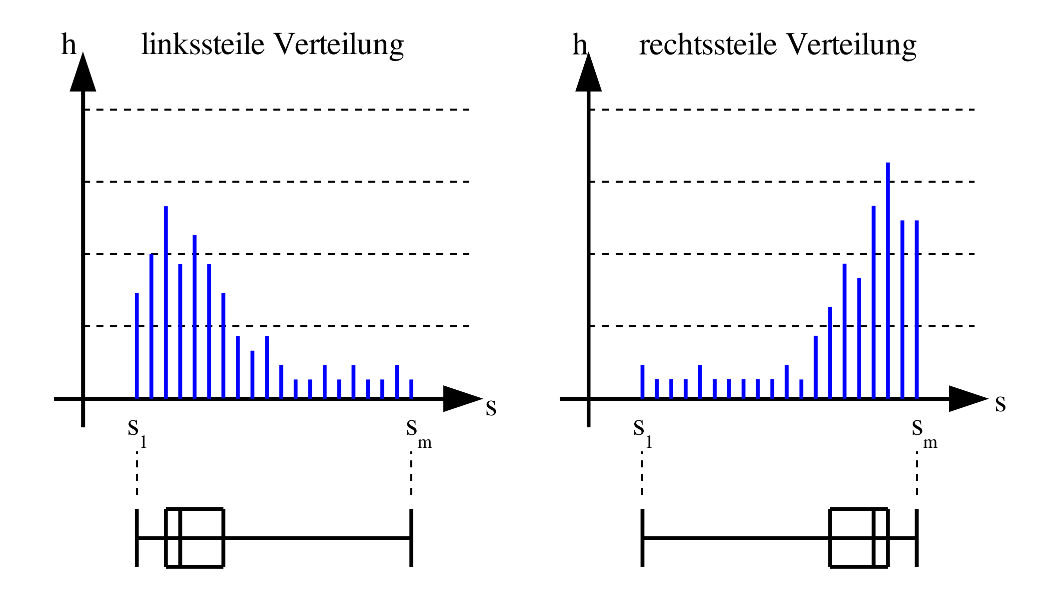
Entsprechend entstehen zwei verschiedene Punkte, je nach der Art des Abstands, den man minimiert. Minimiert man $d_1(x, \alpha)$, so erhält man $\alpha = Z(x)$, minimiert man $d_2(x, \alpha)$, so erhält man $\alpha = A(x)$.

Lageregeln:

Eine einfache grafische Darstellung der Verteilung der Werte x_t bieten die sogenannten **Box-Diagramme**. Wie üblich sortieren wir die Werte x_t aufsteigend der Größe nach: $x_1 \leq x_2 \leq \cdots \leq x_n$. Dann nimmt man den Maximalwert x_n und den Minimalwert x_1 als äussere Begrenzung. Dazwischen trägt man (die Abstände in passendem Verhältnis) das erste und dritte Quartil und den Zentralwert $Z(x) = Q_2(x)$ ein:



Je nachdem ob die Werte x_t häufiger im kleineren oder im größerem Bereich liegen spricht man von einer *links-* bzw. *rechststeilen* Verteilung. Das Box-Diagramm hat in diesen Fällen eine charakteristische Form:



Anhand des Box-Diagrammes in obiger Grafik sieht man sofort, das solche Verteilungen x_t eine besondere Anordnung der Mittelwerte nach sich ziehen. Man spricht von **Lageregeln**, die also zur Definition genommen werden, um zu entscheiden ob eine Verteilung links- oder rechts-steil ist:

$$\begin{array}{rl} \text{linksststeil, wenn} & \text{rechtssteil, wenn} \\ D(x) \leq Z(x) \leq A(x) & A(x) \leq Z(x) \leq D(x) \\ Z(x) < \dfrac{Q_1(x)+Q_3(x)}{2} & \dfrac{Q_1(x)+Q_3(x)}{2} < Z(x) \end{array}$$

Streuungsmaße:

Definition: Nachdem man sich für einen typischen Wert (ein Lagemaß) entschieden hat, fragt man sich natürlich, wie gut dieses die x_t wieder gibt. Anders gesagt: man möchte wissen, wie dicht oder wie weit die x_t um das Lagemaß verstreut sind. Das zugehörige Streuungsmaß soll eben dies leisten. Die einfachste Möglichkeit dies zu tun ist es den **Quartilabstand** Q zu betrachten:

$$Q := Q_3(x) - Q_1(x)$$

Bedenkt man aber, dass $Z(x)$ und $A(x)$ entstanden sind, indem man den zu $x = (x_1, x_2, \ldots, x_n)$ nächst gelegenen Punkt $\alpha = (\alpha, \alpha, \ldots, \alpha)$ auf der Diagonalen genommen hat, so liegt es nahe eben den Abstand zwischen den beiden Punkten x und α als Steuungsmaß zu nehmen. Man definiert daher die **mittlere, absolute Abweichung** $\sigma_1(x)$, die **Varianz** (= mittlere, quadratische Abweichung) $\sigma_2^2(x)$ und die **Standardabweichung** $\sigma_2(x)$:

$$\begin{aligned} \sigma_1(x) &:= \tfrac{1}{n} d_1(x, Z(x)) &&= \frac{1}{n}\sum_{t=1}^{n} |x_t - Z(x)| \\ \sigma_2^2(x) &:= A(x^2) - A(x)^2 &&= \frac{1}{n}\sum_{t=1}^{n} (x_t - A(x))^2 \\ \sigma_2(x) &:= \tfrac{1}{\sqrt{n}} d_2(x, A(x)) &&= \sqrt{\frac{1}{n}\sum_{t=1}^{n} (x_t - A(x))^2} \end{aligned}$$

Man bemerke, dass σ_1 sich auf $Z(x)$ bezieht, da dieser Wert ja den Abstand d_1 minimiert, wohingegen σ_2 sich auf $A(x)$ bezieht, da dies d_2 minimiert. Ausserdem gilt nach Definition $\sigma_2(x)^2 = \sigma_2^2(x)$.

Beispiel: Bei der *Ein-Punkt-Verteilung* haben wir n mal denselben Wert $a \in \mathbb{R}$ vorliegen. D.h. es ist $x_t = a$ für $t \in 1 \dots n$ oder als Datenvektor geschrieben $x = a = (a, a, \dots, a) \in \mathbb{R}^n$. In diesem Fall gilt offenbar $Z(a) = a$ und $A(a) = a$. Und die Streuungsmaße ergeben sich damit zu

$$\begin{aligned} \sigma_1(x) &= \frac{1}{n} \sum_{t=1}^{n} |a - a| = 0 \\ \sigma_2(x) &= \sqrt{\frac{1}{n} \sum_{t=1}^{n} (a - a)^2} = 0 \end{aligned}$$

Beispiel: Bei der *Linearen Verteilung* haben wir nacheinander die Zahlen 1, 2, 3 und so weiter bis n gegeben. D.h. es ist $x_t = t$ für $t \in 1 \dots n$ oder als Datenvektor geschrieben $x = t = (1, 2, 3, \dots, n) \in \mathbb{R}^n$. Ist $n = 2k$ gerade, so gilt offenbar $Z(t) = n/2 = k$ und $A(t) = (n+1)/2$ finden wir für allgemeines n. Die Streuungsmaße ergeben sich damit zu

$$\begin{aligned} \sigma_1(x) &= \frac{1}{n} \sum_{t=1}^{n} |t - k| = \frac{n}{4} \\ \sigma_2(x) &= \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(t - \frac{n+1}{2} \right)^2} = \sqrt{\frac{n^2 - 1}{12}} \end{aligned}$$

Definition: Hat man es mit qualitativen Daten x_t zu tun, hat man als einziges Lagemaß den Modus zur Verfügung. Will man hier ein Streuungsmaß einführen, so muss man sich allein auf die Häufigkeiten stützen. Klar ist: kommt nur s_i vor ($h_i = 1$), so gibt es überhaupt keine Streuung. Und tritt s_i nicht auf ($h_i = 0$), so trägt es zumindest nicht zur Streuung bei. Daher definiert man die **Dispersion** P als:

$$P := \frac{m}{m-1} \sum_{i=1}^{m} h_i(1 - h_i) = \frac{m}{m-1} \left(1 - \sum_{i=1}^{m} h_i^2 \right)$$

Die Dispersion nimmt damit Werte von 0 bis 1 an. Einen Wert von 0 bis $0,8$ interpretieren wir als starke Ballung, Werte von $0,9$ bis 1 als starke Streuung.

Beispiel: Die stärkste Ballung tritt auf, wenn einer alles hat, das bedeutet es ist $h_k = 1$ (für ein $k \in 1 \dots m$) und $h_i = 0$ für $i \neq k$. Dann ist stets $h_i(1 - h_i) = 0$ (für alle $i \in 1 \dots m$) und damit auch $P = 0$. Bei einer völligen *Gleichverteilung* $h_i = 1/m$ (für $i \in 1 \dots m$) erhalten wir hingegen

$$P = \frac{m}{m-1} \left(1 - m \cdot \frac{1}{m^2} \right) = 1$$

Definition: Bei komparativen Daten x_t kann man immerhin noch vergleichen. Man folgt also derselben Idee wie bei der Dispersion, verwendet aber die kumulierten Häufigkeiten H_i anstelle der h_i. Dies führt auf die **Diversität** D, die man definiert, als:

$$D \;:=\; \frac{4}{m-1}\sum_{i=1}^{m} H_i(1-H_i)$$

Dabei erhält man wieder Werte von 0 bis 1, wobei wir Werte von 0 bis 0,6 als starke Ballung, Werte von 0,8 bis 1 als starke Streuung interpretieren.

Beispiel: Bei der *Ein-Punkt-Verteilung* gibt es wieder ein $k \in 1 \dots m$ mit $h_k = 1$ und entsprechend $h_i = 0$ für $i \neq k$. Für die kumulierten Häufigkeiten gilt damit $H_1 = 0, \dots, H_{k-1} = 0, H_k = 1, \dots, H_m = 1$. Und damit $H_i(1 - H_i) = 0$ für alle $i \in 1 \dots m$, mithin $D = 0$. Die größte Streuung erhält man hingegen bei der *Randverteilung*: es kommen nur die beiden äussersten Werte (gleich häufig) vor. Formal bedeutet das $h_1 = 1/2$ und $h_m = 1/2$. und damit $H_1 = 1/2, \dots, H_{m-1} = 1/2, H_m = 1$. Die Diversität wird damit zu

$$D \;=\; \frac{4}{m-1}\left((m-1)\frac{1}{4} + 0\right) \;=\; 1$$

Konzentrationsanalyse:

Satz: Betrachten wir wieder quantitative, positive Daten $x_t \in \mathbb{R}^+$. Wir können bereits beschreiben, wo die Daten liegen und wie weit sie verstreut sind. Nun wollen wir noch ein Maß dafür haben, wie ungleich (ungerecht) die Verteilung ist. Zunächst beobachtet man, dass (sofern $s_1 \leq s_2 \leq \dots \leq s_m$) für alle $k \in 1 \dots m$ gilt:

$$L_k \;\leq\; H_k$$

Definition: Man beachte, dass $L_0 = 0 = H_0$ und $L_m = 1 = H_m$ sind. Wenn wir diese Punkte $(H_k \mid L_k)$ also mit Geradenstücken verbinden, so hängt die Kurve unter der Diagonalen durch. Diese stückweise lineare Funktion mit den Knickstellen $(H_k \mid L_k)$ nennen wir **Lorenz-Kurve** der Verteilung. Es bietet sich an, die durchhängende Fläche der Kurve, als Maß für die Ungleichheit in der Verteilung zu nehmen:

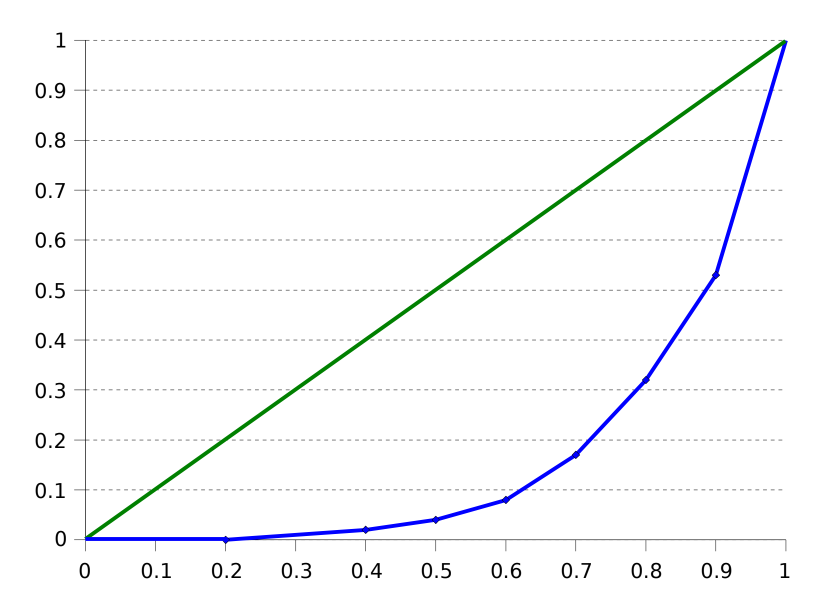
$$\begin{aligned} R \;&:=\; 2\int_0^1 H - L(H)dH \\ &=\; 1 - \sum_{i=1}^{m} h_i(L_{i-1} + L_i) \\ &=\; \frac{2}{nX}\sum_{t=1}^{n} t x_t - \frac{n+1}{n} \end{aligned}$$

Wir nennen R den **Gini - Koeffizienten** der Verteilung. Er nimmt Werte von 0 bis 1 an, wobei wir Werte von 0 bis 0,25 als gute Gleichverteilung und Werte von 0,4 bis 1 als starke Ungleichverteilung interpretieren.

Beispiel: Angenommen in Deutschland gäbe es 10 Haushalte und das Gesamtvermögen der Haushalte sei auf 100 Einheiten normiert. Dann sähe die Vermögensverteilung in Deutschland etwa folgendermaßen aus:

$$0 \leq 0 \leq 1 \leq 1 \leq 2 \leq 4 \leq 9 \leq 15 \leq 21 \leq 47$$

Wir haben also $n = 10$, $m = 8$ und die Zahlenpaare $(H_k \mid L_k)$ betragen hintereinander weg: $(0.2 \mid 0)$, $(0.4 \mid 0.02)$, $(0.5 \mid 0.04)$, $(0.6 \mid 0.08)$, $(0.7 \mid 0.17)$, $(0.8 \mid 0.32)$, $(0.9 \mid 0.53)$ und $(1, 1)$. Der Gini-Koeffizient berechnet sich damit zu $R = 0.67$ und obwohl dieser Wert damit eine sehr starke Ungleichverteilung anzeigt, gehört Deutschland damit zu den eher gemässigten Nationen. Die Lorenz-Kurve illustriert die Ungleichverteilung am eindrucksvollsten:



Lorenz-Kurve (unten) der Vermögensverteilung in Deutschland, die Fläche zwischen den Kurven ist der halbe Gini-Koeffizient R

Beispiel: Bei einer *Gleichverteilung* besitzt jeder gleich viel, d.h. es ist $x_1 = x_2 = \cdots = x_n = a$ konstant. Es kommt also überhaupt nur ein Wert $s_1 = a$ vor und damit sind $m = 1$, $H_1 = 1$ und $L_1 = 1$. Entsprechend ist die Lorenz-Kurve $L(H) = H$ die Diagonale und mithin $R = 0$.

Beispiel: Die schärfste Ungleichverteilung liegt vor, wenn einer alles hat. D.h. es sind $x_1 = x_2 = \cdots = x_{n-1} = 0$ und $x_n = a$. Es gibt also zwei verschiedene Werte $s_1 = 0$ und $s_2 = a$ mit den absoluten Häufigkeiten $n_1 = n-1$ und $n_2 = 1$. Wir finden also die Punkte $(H_1 \mid L_1) = ((n-1)/n \mid 0)$ und $(H_2 \mid L_2) = (1, 1)$. Der Gini-Koeffizient kann damit über die Dreiecksfläche berechnet werden, nach $(1 - R)/2 = 1/2 \cdot 1/n$ und damit $R = 1 - 1/n$. Man kann R also beliebig dicht an 1 rutschen lassen, indem man in dieser Verteilung die Zahl n der Personen erhöht.

Verzerrte Durchschnitte:

Es bezeichne x_t den Kurs einer bestimmten Aktie im Monat t. In einem Ansparmodell werde monatlich der Betrag a in diese Aktie investiert. Die Zahl der im Monat t gekauften Aktien beträgt also a/x_t. Und insgesamt wurden in den Monaten 1 bis n also $a/x_1 + a/x_2 + \cdots + a/x_n$ viele Aktien gekauft. Der durchschnittliche Einkaufspreis $\overline{x}$ pro Aktie beträgt bei dieser Investitionsform also

$$\overline{x} \;=\; \frac{\text{investiertes Geld}}{\text{gekaufte Aktien}} \;=\; \frac{na}{\sum_{t=1}^{n} \frac{a}{x_t}} \;=\; \frac{n}{\sum_{t=1}^{n} \frac{1}{x_t}}$$

Durch diese Betrachtung inspiriert definieren wir das **harmonische Mittel**

$$H(x) \;:=\; \frac{n}{\sum_{t=1}^{n} \frac{1}{x_t}} \;=\; \frac{1}{\sum_{i=1}^{m} h_i \frac{1}{s_i}}$$

Man beachte das das harmonische Mittel also der Kehrwert des arithmetischen Mittels der Kehrwerte ist. In Formeln ist dies viel einfacher:

$$\frac{1}{H(x)} \;=\; \frac{1}{n}\sum_{t=1}^{n} \frac{1}{x_t} \;=\; \sum_{i=1}^{m} h_i \frac{1}{s_i}$$

Sei nun a ein anfänglich investierter Betrag, der im Jahr t mit dem Zinssatz z_t verzinst wird. D.h. im Jahr t wächst der Kontostand mit dem Faktor $x_t := 1 + z_t/100$. Nach einem Jahr ist der Kontostand also ax_1, nach 2 Jahren ax_1x_2 und so weiter, nach n Jahren eben $x_1x_2 \dots x_n a$. Wir fragen uns nun welchem *durchschnittlichen Zinssatz* $\overline{z}$ dies entspricht. Für den durchschnittlichen Wachstumsfaktor $\overline{x}$ gilt offenbar

$$\overline{x}^n a \;=\; x_1x_2 \dots x_n a \quad\Longrightarrow\quad \overline{x} \;=\; \big(x_1x_2 \dots x_n\big)^{\frac{1}{n}}$$

Der durchschnittliche Zinssatz $\overline{z}$ lässt sich aus $\overline{x}$ auch wieder leicht berechnen, als $\overline{z} = 100(\overline{x} - 1)$. Und durch diese Betrachtung inspiriert definieren wir das **geometrische Mittel**

$$G(x) \;:=\; \left(\prod_{t=1}^{n} x_t\right)^{\frac{1}{n}} \;=\; \left(\prod_{i=1}^{m} s_i^{n_i}\right)^{\frac{1}{n}}$$

Analog zum harmonischen Mittel (das ein mit $x \mapsto 1/x$ verzerrtes arithmetisches Mittel ist), ist das geometrische Mittel daher mit dem Logarithmus verzerrt worden. Indem man $x^{1/n} = \exp(\ln(x)/n)$ verwendet führt eine kurze Rechnung auf:

$$\ln(G(x)) \;=\; \frac{1}{n}\sum_{t=1}^{n} \ln(x_t) \;=\; \sum_{i=1}^{m} h_i \ln(s_i)$$

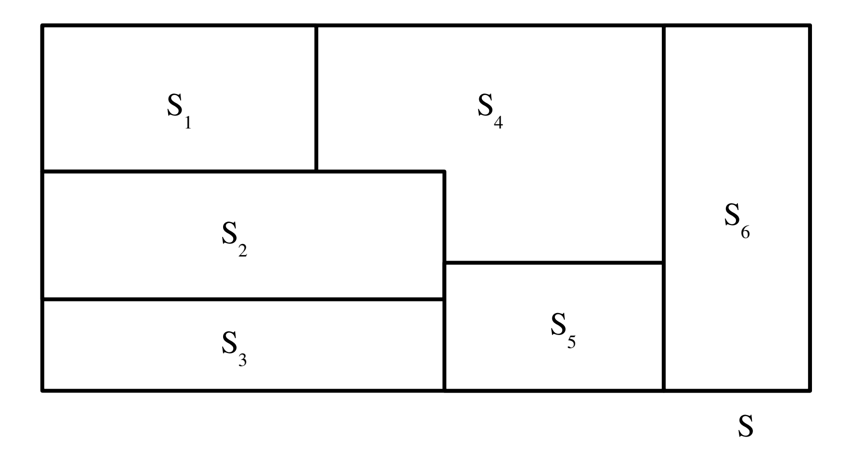
Kapitel 7

Klassierte Verteilungen

Hat man es mit grossen Datenmengen zu tun, kommen oft unsinnig viele Ausprägungen vor. Unterscheiden sich etwa zwei Kontostände nur um 4 Cent, so möchte man diese beiden wie einen Kontostand behandeln. Deswegen **klassiert** man S in diesen Fällen. D.h. man zerlegt S in Teilmengen $S_i \subseteq S$ (mit $i \in 1 \dots m$), formal bedeutet das

(1) $S = S_1 \cup S_2 \cup \cdots \cup S_m$ und

(2) $S_i \cap S_j = \emptyset$ für $i \neq j$



Der Vorteil ist, dass dadurch die Datenmenge übersichtlicher wird. Dies erkauft man sich aber um den Preis, dass man das Wissen über die Verteilung *innerhalb* der Klassen S_i verliert.

Im Normalfall ist $S = [a, b[$ ein Intervall, das zerlegt wird in $S_i = [a_{i-1}, a_i[$ wobei $a = a_0 < a_1 < a_2 < \cdots < a_m = b$. Erhoben wird dann die Zahl n_i der Merkmalsträger, deren Ausprägung x_t in der Klasse S_i liegt. Die relative Häufigkeit h_i und die kumulierten Häufigkeiten lauten dann ganz genau, wie im unklassierten Fall ($i,\, k \in 1 \ldots m$):

$$
\begin{aligned}
n_i &:= \#\{\, t \in 1 \ldots n \mid x_t \in S_i \,\} \\
&= \#\{\, t \in 1 \ldots n \mid a_{i-1} \le x_t \le a_i \,\} \\
h_i &:= \frac{n_i}{n} \\
N_k &:= \sum_{i=1}^{k} n_i \;=\; N_{k-1} + n_k \\
H_k &:= \frac{N_k}{n} \;=\; \sum_{i=1}^{k} h_i \;=\; H_{k-1} + h_k
\end{aligned}
$$

Da man nun aber nicht mehr weiss, wie die Verteilung innerhalb der Klasse S_i aussieht, geht man von folgender *Grundannahme* aus: alle $s \in S_i$ sind gleich häufig. D.h. man geht von der absoluten Häufigkeit h_i zu einer **Häufigkeitsdichte** h_i^* über. Da die Gesamthäufigkeit der Klasse S_i ja h_i sein soll, muss h_i^* also folgenden Wert haben ($i \in 1 \ldots m$):

$$
\begin{aligned}
w_i &:= a_i - a_{i-1} \\
h_i^* &:= \frac{h_i}{w_i} \\
s_i^* &:= \frac{a_{i-1} + a_i}{2}
\end{aligned}
$$

Dabei ist w_i die **Klassenbreite** und s_i^* der **Klassenmittelpunkt**. D.h. s_i^* sagt uns, wo die Werte zu finden sind und S_i ragt links (negativ) und rechts (positiv) um $w_i/2$ über s_i^* heraus. Führt man eine Klassierung von S ein, so muss man also konsequent s_i durch s_i^* ersetzen. Führen wir dies für die Basiswerte aus, so erhalten wir ($i,\, k \in 1 \ldots m$):

$$
\begin{aligned}
X_i^* &:= n_i \int_{a_{i-1}}^{a_i} h_i^* s \, ds \;=\; n_i s_i^* \\
X^* &:= \sum_{i=1}^{m} X_i^* \;=\; \sum_{i=1}^{m} n_i s_i^* \\
A^*(x) &:= \frac{1}{n} X^* \;=\; \sum_{i=1}^{m} h_i s_i^* \\
\ell_i^* &:= \frac{X_i^*}{X^*} \;=\; \frac{h_i s_i^*}{A^*(x)} \\
L_k^* &:= \sum_{i=1}^{k} \ell_i^* \;=\; L_{k-1}^* + \ell_k^*
\end{aligned}
$$

Beispiel: In einem Unternehmen werden die folgenden Gehälter s_i gezahlt:

i	Gehalt s_i	Häufigleit n_i	in Klasse
1	$0,64$	4	1
2	$0,9$	1	1
3	1	2	1
4	$1,4$	1	1
5	$1,8$	2	1
6	2	4	2
7	$3,5$	3	2
8	$5,5$	1	2
9	9	1	3
10	$9,5$	1	3

Unter den $n = 20$ Mitarbeitern gibt es also $m = 10$ verschiedene Gehälter, eigentlich noch kein Anlass zum klassieren, wir tun es aber dennoch und führen 3 verschiedene Klassen ein: die Geringverdiener $S_1 = [0, 2[$, den Mittelstand $S_2 = [2, 6[$ und die Großverdiener $S_3 = [6, 10[$. Es ist dann

S_i	n_i	w_i	h_i	h_i^*	s_i^*	X_i^*
$[0, 2[$	10	2	$0,5$	$0,25$	1	10
$[2, 6[$	8	4	$0,4$	$0,1$	4	32
$[6, 10[$	2	4	$0,1$	$0,025$	8	16

Das Gesamtsumme der Gehälter wird nach der Klassierung also auf $X^* = 58$ geschätzt. In Wirklichkeit beträgt die Gesamtsumme aber $X = 52,96$. Die Abweichung kommt daher, dass die Verteilung in den Klassen eben nicht gleichmässig ist - vor allem in Klasse 3 besteht eine große Abweichung.

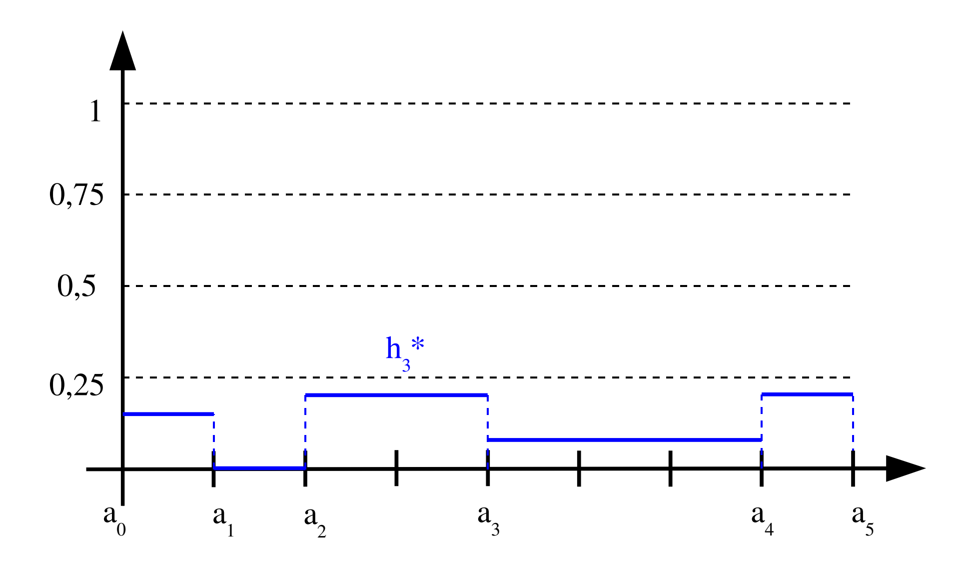
Lagemaße:

Analog zum unklassierten Fall, führt man die Lagemaße ein. Der **Modus** (oder auch *dichtester Wert*) $D^*(x)$ ist der Wert mit der größten Häufigkeitsdichte. Das **arithmetische Mittel** (oder auch *Durchschnitt*) $A^*(x)$ ist der Gesamtbetrag X^* pro Merkmalsträger, also:

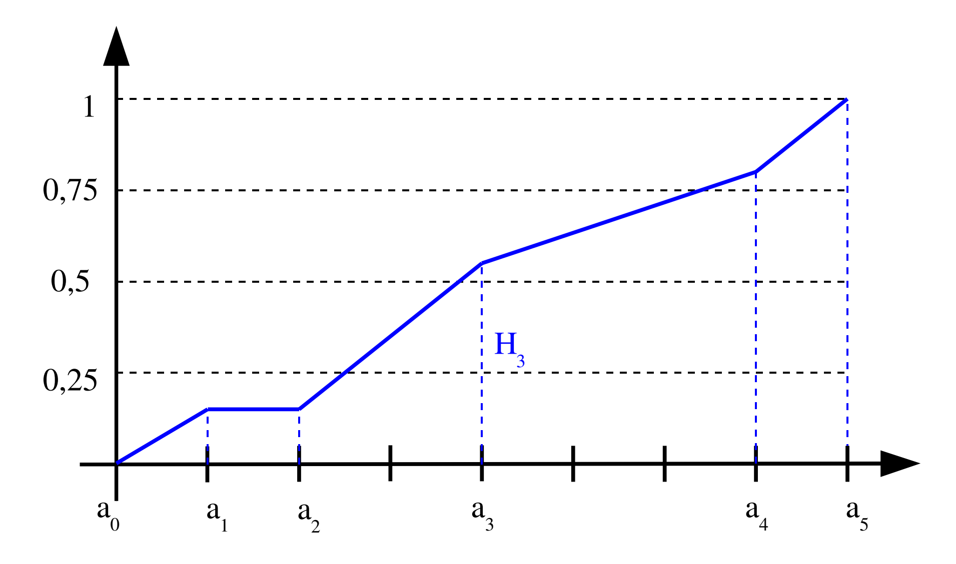
$$\begin{aligned} D^*(x) &:= s_k^* \quad \text{so dass} \quad h_k^* = \max\{\, h_i^* \mid i \in 1 \dots m \,\} \\ A^*(x) &:= \frac{1}{n} X^* \quad = \quad \sum_{i=1}^{m} h_i s_i^* \end{aligned}$$

In einem **Histogramm** werden die Häufigkeiten graphisch dargestellt, indem man die Punkte $(a_{i-i} \mid h_i^*)$ und $(a_i \mid h_i^*)$ zu Balken verbindet. Der Modus ist also die Klasse, mit dem höchsten Balken. Verbindet man hingegen die Punkte $(a_i \mid H_i)$ zu einer stückweise linearen Funktion, so spricht man von der **empirischen Verteilungsfunktion** H.

Beispiel: Wir klassieren das Intervall $[0, 8[$ an den Stellen $a_0 = 0$, $a_1 = 1$, $a_2 = 2$, $a_3 = 4$, $a_4 = 7$ und $a_5 = 8$. Die relativen Häufigkeiten der Klassen seien $h_1 = 0.15$, $h_2 = 0$, $h_3 = 0.4$, $h_5 = 0.25$ und $h_5 = 0.2$. Dann erhalten wir daraus die Häufigkeitsdichten $h_1^* = 0.15$, $h_2^* = 0$, $h_3^* = 0.2$, $h_4^* \approx 0.08$ und $h_5^* = 0.2$. In diesem Fall ist der dichteste Wert also nicht eindeutig! Er beträgt $s_3^* = 3$ oder auch $s_5^* = 7.5$. Das Histogramm bzw. die empirische Verteilungsfunktion zu diesen Daten sehen damit wie folgt aus:



Histogramm



Verteilungsfunktion

Die Stelle $\overline{x}$, an der diese Funktion H den Wert $1/2$ annimmt, wird **Median** (oder *Zentralwert*) genannt (es ist $k \in 1 \ldots m$ so, dass $H_{k-1} < 1/2 \leq H_k$):

$$\begin{aligned} Z^*(x) &:= \overline{x} \quad \text{so dass} \quad H(\overline{x}) = \frac{1}{2} \\ &= a_k - \frac{H_k - \frac{1}{2}}{h_k^*} \end{aligned}$$

Genau wie im unklassierten Fall kann man nicht nur die Trennlinien zwischen unterer und oberer Hälfte markieren, sondern auch zwischen den jeweiligen Vierteln. D.h. man führt wiederum **Quartile** ein, wobei die Trennlinie jetzt innerhalb einer Klasse liegen kann. Ganz analog zum Zentralwert setzt man:

$$\begin{aligned} Q_j^*(x) &:= \overline{x} \quad \text{so dass} \quad H(\overline{x}) = \frac{j}{4} \\ &= a_k - \frac{H_k - \frac{j}{4}}{h_k^*} \end{aligned}$$

Streuung und Konzentration:

Man kann die *mittlere, absolute Abweichung*, *Varianz*, *Standardabweichung* und sogar den *Gini-Koeffizienten* genauso im klassierten Fall berechnen, wie im unklassierten Fall. Aufgrund der Grundannahme treten aber zu der *externen* Streuung zwischen den Klassen noch die *internen* Streuungen innerhalb der Klassen hinzu. Man erhält also Korrekturterme, die die interne Streuung widerspiegeln:

$$\begin{aligned} \sigma_2^*(x)^2 &:= \sum_{i=1}^{m} h_i^* \int_{a_{i-1}}^{a_i} (s - A^*(x))^2 \, ds \\ &= \sum_{i=1}^{m} h_i (s_i^* - A^*(x))^2 + \frac{1}{12} \sum_{i=1}^{m} h_i w_i^2 \\ \sigma_1^*(x) &:= \sum_{i=1}^{m} h_i^* \int_{a_{i-1}}^{a_i} |s - Z^*(x)| \, ds \\ &= \sum_{i=1}^{m} h_i |s_i^* - Z^*(x)| + h_k^* \left(\frac{w_k}{2} - |s_k^* - Z^*(x)| \right)^2 \end{aligned}$$

wobei in letzterer Gleichung $k \in 1 \ldots m$ so war, dass $a_{k-1} \leq Z^*(x) \leq a_k$. Genauso erfährt der Gini-Koeffizient eine Abweichung durch die Ungleichverteilung innerhalb der Klassen:

$$R^* = 1 - \sum_{i=1}^{m} h_i (L_{i-1}^* + L_i^*) + \frac{1}{6A^*(x)} \sum_{i=1}^{m} h_i^2 w_i$$

Kapitel 8

Zweidimensionale Verteilungen

Wir haben bisher immer nur ein Merkmal x betrachtet - wo es liegt, wie weit es streut und sogar wie ungleich es verteilt ist. Oftmals interessiert man sich aber gerade für den Zusammenhang zwischen zwei Merkmalen x und y. Entsprechend haben wir diesmal also zwei Mengen R und S möglicher Merkmalsausprägungen. Wir bezeichnen

$$\begin{aligned} R &= \{ r_1, r_2, \ldots, r_p \} \\ S &= \{ s_1, s_2, \ldots, s_q \} \end{aligned}$$

Einem jeden Merkmalstrğer $t \in 1 \ldots n$ werden also zwei Merkmale x_t und y_t zugeordnet. D.h. x und y sind zwei Zuordnungen der Form

$$\begin{aligned} x &: 1 \ldots n \to R : t \mapsto x_t \\ y &: 1 \ldots n \to S : t \mapsto y_t \end{aligned}$$

Wir haben es also mit einem Merkmalsraum zu tun, der die Paare (x_t, y_t) enthält. Und als diesen bietet sich das *Karthesische Produkt* an:

$$R \times S = \{ (r_i, s_j) \mid i \in 1 \ldots p \text{ und } j \in 1 \ldots q \}$$

Wir definieren nun die Grundbegriffe, wie im eindimensionalen Fall. Wir müssen lediglich der Tatsache Rechnung tragen, dass wir immer die Paare (x_t, y_t) betrachten. Die *absolute Häufigkeit* $n_{i,j}$ bzw. *relative Häufigkeit* $h_{i,j}$ gibt also die Zahl der Vorkommen *der Kombination* (r_i, s_j) an:

$$\begin{aligned} n_{i,j} &:= \# \{ t \in 1 \ldots n \mid x_t = r_i \text{ und } y_t = s_j \} \\ h_{i,j} &:= \frac{n_{i,j}}{n} \end{aligned}$$

Am einfachsten lassen sich diese Daten in einer Matrix (Tabelle) darstellen. Als Beispiel betrachten wir eine repräsentative Umfrage unter 1000 Bundesbürgern zu ihrer Berufsausbildung (Merkmal x mit den Ausprğungen 'keine', 'Lehre' und 'Studium') und ihrem Arbeitsverhältnis (Merkmal y mit den Ausprägungen 'arbeitslos', 'Arbeiter', 'Angestellter' und 'selbstständig'). Das Ergebnis der Umfrage könnte dann wie folgt aussehen:

	arbeitslos	Arbeiter	Angestellter	selbstständig	Summe
keine	35	109	88	16	248
Lehre	47	189	263	36	535
Studium	9	15	152	41	217
Summe	91	313	503	93	1000

Die Zahl aller Merkmalsträger t mit Ausprägung r_i (bzw. mit Ausprägung s_j) bezeichnen wir als **absolute Randhäufigkeit**. Sie ist definiert als:

$$n_{i,+} := \#\{\, t \in 1 \ldots n \mid x_t = r_i \,\} = \sum_{j=1}^{q} n_{i,j}$$

$$n_{+,j} := \#\{\, t \in 1 \ldots n \mid y_t = s_j \,\} = \sum_{i=1}^{p} n_{i,j}$$

Entsprechend definieren wir die **relativen Randhäufigkeiten** als die Anteile der absoluten Randhäufigkeiten an der Gesamtzahl der Merkmalsträger:

$$h_{i,+} := \frac{n_{i,+}}{n} = \sum_{j=1}^{q} h_{i,j}$$

$$h_{+,j} := \frac{n_{+,j}}{n} = \sum_{i=1}^{p} h_{i,j}$$

Wollen wir die Arbeitslosenquote der Akademiker bestimmen, so müssen wir offenbar die Zahl der arbeitslosen Akademiker durch die Gesamtzahl der Akademiker teilen ($= 9/217 = 4,1\%$). Wir sortieren also erst diejenigen aus, die der Bedingung genügen Akademiker zu sein. Und unter diesen betrachten wir dann diejenigen, die auch die Ausprägung 'arbeitslos' aufweisen. Allgemein definieren wir die **bedingte Häufigkeit** $h(r_i \mid s_j)$ als den Anteil der Merkmalsträger t mit Ausprägung (r_i, s_j) an all denen, die die Bedingung s_j erfüllen:

$$h(r_i \mid s_j) := \frac{n_{i,j}}{n_{+,j}} = \frac{h_{i,j}}{h_{+,j}}$$

$$h(s_j \mid r_i) := \frac{n_{i,j}}{n_{i,+}} = \frac{h_{i,j}}{h_{i,+}}$$

In unserem Fall unterscheidet sich die Arbeitslosenquote der Akademiker ($4,1\%$) von der allgemeinen Arbeitslosenquote ($9,1\%$). Es überrascht nicht, dass die Qualifikation das Beschäftigungsverhältnis beeinflusst. Betrachten wir noch eine andere Zahl: etwa ein Drittel (genau 313/1000) aller Bundesbürger sind Arbeiter. Und ein Viertel (genau 248/1000) aller Bundesbürger haben keine Berufsausbildung. Wären diese beiden Eigenschaften voneinander unabhängig, würden sie sich einfach überlagern: ein Viertel der Arbeiter hätte keine Berufsausbildung, das wäre etwa ein Zwölftel (genau $77,624/1000$) der Bundesbürger. Tatsächlich sind es aber 109/1000. Die beiden Merkmale sind also nicht unabhängig. Nun allgemein:

Die Merkmale x und y heißen **unabhängig verteilt**, wenn die bedingten Häufigkeiten gar nicht von der Bedingung abhängen. Formal: die folgenden drei Aussagen (a), (b) und (c) sind äquivalent. Und sind sie erfüllt (es genügt wenn eine dies ist), dann nennen wir x und y unabhängig:

(a) für alle $i \in 1 \dots p$ und alle $j \in 1 \dots q$ gilt: $h_{i,j} \;=\; h_{+,j} h_{i,+}$

(b) für alle $i \in 1 \dots p$ und alle $j \in 1 \dots q$ gilt: $h(r_i \mid s_j) \;=\; h_{i,+}$

(c) für alle $i \in 1 \dots p$ und alle $j \in 1 \dots q$ gilt: $h(s_j \mid r_i) \;=\; h_{+,j}$

Diese Definition ist rein qualitativ - die Merkmale x und y sind unabhängig ja oder nein. Wir suchen nun nach einem Kriterium wie stark die beiden Merkmale voneinander abhängen. Dazu bezeichnen wir die Idealwert der Unabhängigkeit mit:

$$u_{i,j} \;=\; h_{+,j} h_{i,+}$$

Diese Werte bilden wieder eine Häufigkeitsverteilung - nämlich die Verteilung, die durch unabhängige Überlagerung der Verteilung $h_{i,+}$ von R und $h_{+,j}$ von S entstanden wäre. Entsprechend gilt wieder:

$$\sum_{i=1}^{p} \sum_{j=1}^{q} u_{i,j} \;=\; \sum_{i=1}^{p} h_{i,+} \sum_{j=1}^{q} h_{+,j} \;=\; \sum_{i=1}^{p} h_{i,+} \;=\; 1$$

Dann messen wir die quadratische Abweichung der Häufigkeit $h_{i,j}$ von dem Idealwert $u_{i,j}$ der Unabhängigkeit. Damit aber alle Kombinationen (i,j) gleich stark berücksichtigt werden, müssen wir die Abweichung wieder mit $u_{i,j}$ normieren. Auf diese Weise findet man die **mittlere quadratische Kontingenz**:

$$C \;:=\; \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(h_{i,j} - u_{i,j})^2}{u_{i,j}}$$

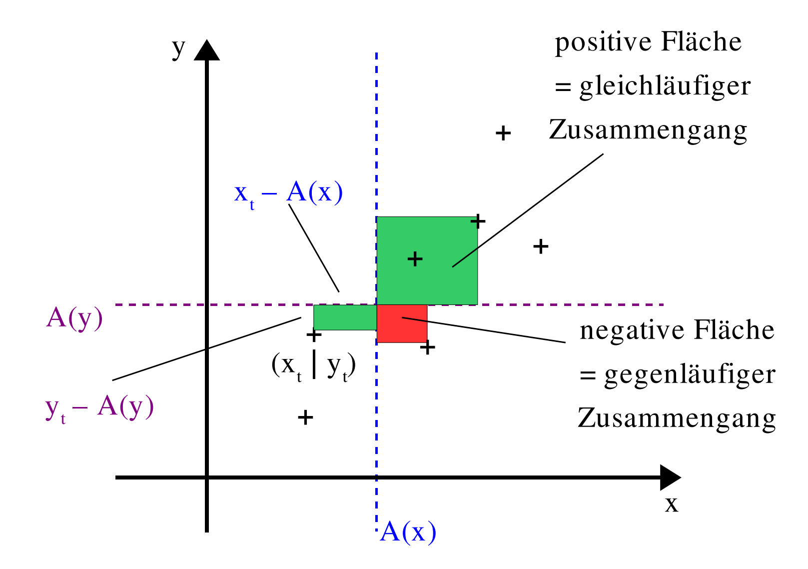
Das so definierte C nimmt Werte von 0 bis $\mu - 1$ an, wobei $\mu := \min\{\, p, q \,\}$. Der Wert 0 wird dann und nur dann angenommen, wenn x und y unabhängig verteilt sind. Und der Wert μ wird erreicht, wenn pro Zeile (oder pro Spalte) immer nur ein $h_{i,j} \neq 0$ auftritt. Wir definieren daher die **korrigierte Kontingenz**:

$$C^* \;:=\; \sqrt{\frac{\mu}{\mu - 1}} \sqrt{\frac{C}{C + 1}} \;\in\; [0, 1]$$

Kapitel 9

Korrelation

Wir betrachten wieder zwei Merkmale x und y und wollen diesmal den Zusammenhang genauer analysieren. Zusammenhang heißt, dass besondere (große oder kleine) y_t auch zu besonderen x_t gehören müssen. Sind die besonderen y_t wirr verteilt, haben die Daten wohl auch nichts miteinander zu tun. Ob ein Wert x_t oder y_t aber groß oder klein ist, messen wir daran, wie weit er vom Durchschnitt $A(x)$ bzw. $A(y)$ abweicht. Sind beide Abweichungen groß spricht das für einen starken Zusammenhang. Ist eine der beiden Abweichungen groß die andere klein, spricht das für einen schlechten Zusammenhang. Die Kombination der Abweichungen kann man durch die Produkte $(x_t - A(x))(y_t - A(y))$ als Fläche im xy-Diagramm messen.



Man beachte, dass klein-klein und groß-groß positive Produkte bilden, während klein-groß und groß-klein negative Produkte bewirken. Sind die Werte wirr verteilt, heben sich die Flächen also gegenseitig weg. Steckt ein System dahinter, addieren sie sich auf. Ausgehend von dieser Überlegung definiert man die **Kovarianz**:

$$\begin{aligned}\sigma(x \mid y) &:= \frac{1}{n}\sum_{t=1}^{n}(x_t - A(x))(y_t - A(y)) \\ &= \sum_{i=1}^{p}\sum_{j=1}^{q} h_{i,j}(r_i - A(x))(s_j - A(y))) \\ &= A(xy) - A(x)A(y)\end{aligned}$$

Satz: Anhand dieser Definition rechnet man leicht nach, dass die Kovarianz eine Reihe nützlicher Eigenschaften erfüllt ($a \in \mathbb{R}$, x, x' und $y \in \mathbb{R}^n$):

$$\begin{aligned}\sigma(x \mid y) &= \sigma(y \mid x) \\ \sigma(x \mid x) &= \sigma(x)^2 > 0 \\ \sigma(ax \mid y) &= a\sigma(x \mid y) \\ \sigma(x + x' \mid y) &= \sigma(x \mid y) + \sigma(x' \mid y)\end{aligned}$$

Nun hat die Kovarianz noch einen Webfehler: sie misst die absolute Abweichung der Werte x_t und y_t von den Durchschnittswerten. Es macht aber mehr Sinn diese Abweichung in Einheiten der jeweiligen Standardabweichung auszudrücken. Dies führt zur Definition des **Korrelationskoeffizienten**:

$$\begin{aligned}r &:= \frac{1}{n}\sum_{t=1}^{n}\frac{x_t - A(x)}{\sigma(x)} \cdot \frac{y_t - A(y)}{\sigma(y)} \\ &= \frac{\sigma(x \mid y)}{\sigma(x)\sigma(y)} = \frac{\langle x - A(x) \mid y - A(y)\rangle}{\|x - A(x)\| \cdot \|y - A(y)\|} \\ &= \cos\left(\begin{array}{c}\text{Schnittwinkel von} \\ x - A(x) \text{ und } y - A(y)\end{array}\right)\end{aligned}$$

Ein Korrelationskoeffizient von -1 bis $-0,6$ zeigt einen starken, gegenläufigen, linearen Zusammenhang an. Ein Korrelationskoeffizient von $-0,4$ bis $0,4$ zeigt an, dass kein ausgeprägter linearer Zusammenhang besteht. Und $0,6$ bis 1 zeigt einen starken, gleichläufigen, linearen Zusammenhang an.

Lineare Regression:

Wir vermuten nun einen Zusammenhang $y = f(x)$ zwischen den Größen x und y. Die Funktion f, die den Zusammenhang beschreiben soll hängt aber zusätzlich von gewissen Parametern a, b, $c, \dots$ ab. Wir werden uns hier aber nur mit dem Fall eines linearen Zusammenhangs beschäftigen:

$$f_{a,b}(x) = ax + b$$

Satz: Unsere Aufgabe besteht also darin, die best möglichen Werte für die Parameter a und b zu finden, so dass die Funktion sich möglichtst dicht an die Datenpunkte (x_t, y_t) anschmiegt. Sich möglichst dicht anzuschmiegen bedeutet formal, dass der Abstand

$$d_2(a,b) = \sqrt{\sum_{t=1}^{n}(ax_t + b - y_t)^2}$$

minimal wird. Löst man dieses Minimierungsproblem, so findet man (durch Ableitung von $d_2(a,b)$ nach b bzw. nach a) das folgende Gleichungssystem

$$\begin{aligned} A(x)a + b &= A(y) \\ A(x^2)a + A(x)^2 b &= A(xy) \end{aligned}$$

Löst man dieses Gleichungssystem nach a und b auf und vergleicht die entstehenden Ausdrücke mit der Definition der Varianz bzw. Kovarianz, so erkennt man die folgenden Berechnungsformeln für a und b:

$$\begin{aligned} a &= \frac{\sigma(x \mid y)}{\sigma(x)^2} \\ b &= A(y) - aA(x) \end{aligned}$$

Diese Gerade $y = ax + b$ heißt **Ausgleichsgerade** oder auch **Regressionsgerade** von x und y. Bemerkenswert ist, dass diese Gerade (nach Gleichung 1 des Minimierungsproblems) durch den Punkt $(A(x), A(y))$ der arithmetischen Mittel läuft. Und damit gilt:

$$A(ax + b) = aA(x) + b = A(y)$$

Bezeichnen wir die Abweichung der y_t von der Regressionsgeraden mit $u_t := y_t - (ax_t + b)$ so ist also $A(u) = 0$ und damit $\sqrt{n}\sigma(u) = d_2(a,b)$. Wir haben also die Gerade so gelegt, dass sie die Streuung der Abweichungen minimiert.

Man kann in eine Wolke von Datenpunkten (x_t, y_t) natürlich nicht nur Geraden einfitten, sondern beliebige andere Funktionen. Wie die Berechnung der Regressionsfunktion für Polynome höheren Grades funktioniert findet sich in den mathematischen Ergänzungen (16.15.)

Bisher haben wir nur eine Gerade durch die Datenpunkte (x_t, y_t) gelegt, dabei diente x als Basis und y war der zugeordnete Wert. Entsprechend bezeichnen wir diese Gerade mit $g_x(x) = a_x x + b_x$. Natürlich können x und y aber auch die Rollen tauschen, d.h. y wird zur Basis und $g_y(y) = a_y y + b_y$ ist der zugeordnete Wert. Die Formeln für a_y und b_y sind (durch Vertauschen von x und y) $a_y = \sigma(x \mid y)/\sigma(y)^2$ und $b_y = A(x) - a_y A(y)$. Damit folgt dann

$$\begin{aligned} a_x &= \frac{\sigma(y)}{\sigma(x)} r \\ a_x a_y &= r^2 \end{aligned}$$

Aus diesen beiden Formeln folgt: (1) $r = \pm 1$ tritt genau dann auf, wenn die beiden Geraden g_x und g_y gleich sind und (2) $r = 0$ tritt genau dann auf, wenn die Geraden g_x und g_y senkrecht zueinander stehen. Schließlich gilt auch noch die **Streuungszerlegung**

$$\sigma(y)^2 = \sigma(a_x x + b_x)^2 + \sigma(u)^2$$

Dies ist so zu interpretieren: die Streuung der y_t besteht aus 2 Anteilen (1) der durch den linearen Zusammenhang $y = a_x x + b_x$ vermittelten Streuung der x_t und (2) der Reststreuung $\sigma(u)^2$, die wir nicht erklären können. Der Anteil (an der Gesamtstreuung der y_t) der durch den linearen Zusammenhang $y_t = a_x x_t + b_x$ und die Streuung der x_t erklärbar ist, ist damit gegeben durch:

$$\frac{\sigma(a_x x + b_x)^2}{\sigma(y)^2} = r^2$$

Der Regressionskoeffizient liefert also dreierlei: (1) ein Maß dafür, wie ausgeprägt der lineare Zusammenhang ist (stark ab einem Betrag von $0,6$), (2) r ist positiv, wenn die Daten gleichsinnig verlaufen und negativ, wenn die Daten gegensinnig verlaufen und (3) $1 - r^2$ gibt an, wieviel Prozent der Streuung der y_t nicht durch den linearen Zusammenhang erklärt werden kann. Doch Vorsicht: ist $r = 0$ so heißst das *nicht*, dass kein Zusammenhang zwischen den Daten besteht. Es besteht nur kein *linearer* Zusammenhang!

Rangkorrelation:

Die bisherige Berechnung der Korrelation und der Ausgleichsgeraden funktioniert nur im Fall von quantitativer Daten $S \subseteq \mathbb{R}$. Ist eines der Merkmale x oder y nur komparativ, muss man eine neue Idee haben. Oder man kocht einfach dieselbe Idee nochmal auf: man weist den komparativen Daten einfach einen numerischen Wert (genannt *Rang*) zu, der ihre Reihenfolge wieder gibt. Das einfachste wäre es, die Daten x_t zu sortieren $x_1 \leq x_2 \leq \cdots \leq x_n$ und der Reihe nach mit 1, 2 bis n durch zu nummerieren: $R(x_t) := t$. Das Problem ist, das manche x_t aber mehrfach vorkommen und *eine* Ausprägung x_t ja nicht *mehrere* verschiedene Rangzahlen $R(x_t)$ haben kann. In diesem Fall bildet man also das arithmetische Mittel dieser Zahlen.

Beispiel: Wir betrachten die Berufsausbildung von 6 Personen und sortieren sie aufsteigend: keine, Lehre, Lehre, Lehre, Studium, Studium. Die Ausprägung 'keine' ist unkritisch, sie kommt an erster Stelle und erhält den Rang $R(\text{keine}) = 1$. Die Ausprägung 'Lehre' kommt aber 3mal vor (an den Stellen 2, 3 und 4) und erhält damit den Rang $R(\text{Lehre}) = (2+3+4)/3 = 3$. Auch die Ausprägung 'Studium' kommt mehrfachl vor (an den Stellen 5 und 6), erhält also den Rang $R(\text{Studium}) = (5+6)/2 = 5.5$.

Satz: Allgemein kann man den Rang der Ausprägung s_i auch direkt aus den kumulierten, absoluten Häufigkeiten N_i berechnen, nach der Formel

$$R(s_i) \quad = \quad \frac{1}{n_i}\Big(N_{i-1} + 1 + \cdots + N_{i-1} + n_i\Big) \quad = \quad N_{i-1} + \frac{n_i + 1}{2}$$

Ersetzt man jedes x_t also durch seine Rangzahl $R(x_t)$ und verfährt genauso mit dem zweiten Merkmal y, so kann man wieder den Korrelationskoeffizienten r für die Ränge bestimmen. Prinzipiell könnte man auch die Ausgleichsgerade $R(y) \approx aR(x) + b$ bestimmen, es ist aber unklar, wie diese zu interpretieren ist. Bemerkenswert ist aber auch, dass das arithmetische Mittel der Rangzahlen stets $(n+1)/2$ ist

$$A(R(x)) \quad = \quad \frac{n+1}{2}$$

Kapitel 10

Zeitreihenanalyse

Die Zeitreihenanalyse ist eine leichte Modifikation der linearen Regression. Das Problem sieht folgendermaßen aus: zu jedem Zeitpunkt $t \in 1 \ldots n$ hat man einen Wert y_t gegeben. Zumeist ist y_t der Umsatz eines Unternehmens im Quartal t. Der Umsatz ist jedoch von der speziellen Saison (etwa Frühjahr, Sommer, Herbst oder Winter) abhängig. Um den Umsatz für die Zukunft zu prognostizieren, genügt eine lineare Regression also nicht, da man den Einfluss der jeweiligen Saison berücksichtigen muss. Wir gehen im folgenden von der Annahme aus, dass y_t also in 3 Teile zerfällt: (1) einen linearen Trend $g(t) = at + b$, (2) einen periodischen, saisonalen Einfluss s_t und (3) in eine zuflige Störung u_t. Insgesamt also:

$$y_t \;=\; (at + b) + s_t + u_t$$

Wenn wir davon ausgehen, dass eine Periode k Saisons umfasst, erstrecken sich unsere Daten y_t über $m = n/k$ Perioden (Beispiel: bei Quartalen beträgt $k = 4$, bei Daten über $n = 12$ Quartale, haben wir also $m = 3$ Jahre vorliegen). Dass die saisonalen Einflüsse periodisch sind heißt nun

$$\begin{array}{ccccccc} s_1 & = & s_{1+k} & = & s_{1+2k} & = & \ldots \\ s_2 & = & s_{2+k} & = & s_{2+2k} & = & \ldots \\ \vdots & = & \vdots & = & \vdots & = & \ldots \\ s_k & = & s_{2k} & = & s_{3k} & = & \ldots \end{array}$$

Die Zahlenfolge $(s_1, s_2, \ldots, s_k)$ heißt auch **Saisonfigur**. Wir lösen das Problem nun in 2 Schritten: zunächst ermitteln wir den linearen Trend, indem wir eine lineare Regression der Punkte (t, y_t) ausführen. Eine leichte Rechnung zeigt $A(t) = (n + 1)/2$ und damit auch $\sigma(t)^2 = (n^2 - 1)/12$. Mit Hilfe dieser Gleichungen wird die Berechnung der Trendgeraden recht einfach:

$$\begin{aligned} a &= \frac{\sigma(t \mid y)}{\sigma(t)^2} \\ &= \frac{12}{n^2 - 1} A(ty) - \frac{6}{n - 1} A(y) \\ b &= A(y) - aA(t) \\ &= A(y) - \frac{n + 1}{2} a \end{aligned}$$

Die saisonalen Einfüsse berechnen wir, als durchschnittliche Abweichung der y_t von dem erwarteten Wert auf der Trendgeraden. Man beachte, dass nur s_1, s_2 bis s_k berechnet werden müssen, da sich die saisonalen Abweichungen danach zyklisch wiederholen:

$$\begin{aligned} s_i &= \frac{1}{m}\sum_{j=0}^{m-1}(y_{i+jk} - g(i+jk)) \\ &= \frac{1}{m}\sum_{j=0}^{m-1} y_{i+jk} - \left(g(i) + \frac{n-k}{2}a\right) \end{aligned}$$

Sind a, b, s_1, s_2 bis s_k auf diese Weise berechnet worden, so hat man also eine Näherung $at + b + s_t$ für y_t gefunden. Als Prognose für die nächsten Perioden $t > n$ dient damit ebenso $at + b + s_t$. Es bleibt die Frage, wie gut die Zeitreihe die tatsächlichen Verhältnisse wieder gibt. Zunächst beobachtet man, dass die Varianz der y_t in folgende 3 Teile zerfällt:

$$\sigma(y)^2 = \sigma(at+b)^2 + \sigma(s)^2 + \sigma(u)^2$$

Dabei ist $\sigma(at+b)^2$ die Streuung, die aufgrund der Streuung der t (durch den linearen Zusammenhang) auf y_t übertragen wird. Und $\sigma(s)^2$ ist die Streuung der Saisonfigur. Diese beiden Anteile werden durch die Zeitreihenanalyse erklärt. Man beachte, dass die saisonalen Einflüsse sich im Mittel aufheben $A(s) = 0$. Damit kann man diese beiden Anteile recht leicht berechnen:

$$\begin{aligned} \sigma(at+b)^2 &= \frac{n^2-1}{12}a^2 \\ \sigma(s)^2 &= \frac{1}{k}\sum_{i=1}^{k} s_i^2 \end{aligned}$$

Die Reststreuung $\sigma(u)^2$ kann man durch die Zeitreihenanalyse nicht erklären. Wie zuvor bei der linearen Regression nimmt man den erklärten Anteil der Streuung als Maß für die Güte der Näherung. Bei einem Wert r^2 bis $1/3$ ist die Zeitreihenanalyse unpassend, bei einem Wert von $2/3$ oder mehr, gibt sie die Verhältnisse gut wieder.

$$r^2 := \frac{\sigma(at+b)^2 + \sigma(s)^2}{\sigma(y)^2}$$

Eine etwas genauere Betrachtung der Situation ergibt, dass die Aufteilung in 2 Schritte nur zu einer Näherungslösung des Problems führt. Genauere Lösungsformeln, finden sich in den mathematischen Ergänzungen (16.19).

Kapitel 11

Indices

Indices dienen dazu den zeitlichen Verlauf einer Größe zu dokumentieren. Etwa beim Sozialprodukt wird der Gesamtwert aller produzierten Waren aufsummiert. Der Index ist nun das Verhältnis des Sozialprodukts im *Berichtsjahr* b zum Sozialprodukt des *Bezugsjahres* a. Wir befinden uns also in folgender Situation: wir betrachten n verschiedene Waren (die Merkmalsträger), die wir mit $i \in 1 \ldots n$ durchnummerieren. Zu jeder dieser Waren i betrachten wir den Preis $p_{i,t}$ und die Menge $q_{i,t}$ zum Zeitpunkt $t \in \mathbb{N}$. Der Gesamtwert aller dieser Waren summieren sich also zu:

$$\text{Gesamtwert} = \sum_{i=1}^{n} p_{i,t}\, q_{i,t}$$

Die Zahlenfolge $(q_{1,t}, q_{2,t}, \ldots, q_{n,t})$ heißt **Warenkorb** (oder auch Mengengerüst) zum Zeitpunkt t. Den Gesamtwert des Warenkorbs zum Zeitpunkt t berechnet mit den Preisen vom Zeitpunkt s bezeichnen wir mit:

$$M(s \mid t) = \sum_{i=1}^{n} p_{i,s}\, q_{i,t}$$

Die relative Gesamtwertentwicklung vom Zeitpunkt a bis zum Zeitpunkt b bezeichnen wir mit $W_{a,b}$. Offenbar berechnet sie sich nach:

$$W_{a,b} := \frac{M(b \mid b)}{M(a \mid a)}$$

Der *Preisindex nach Laspeyres* bezeichnet die relative Gesamtwertentwicklung wenn man im Berichtsjahr noch immer denselben Warenkorb gehabt hätte, wie im Bezugsjahr. Er ist also wie folgt definiert:

$$P^{L}_{a,b} := \frac{M(b \mid a)}{M(a \mid a)}$$

Der *Preisindex nach Paasche* kehrt das ganze um: er ist die relative Gesamtwertentwicklung wenn man im Bezugsjahr schon denselben Warenkorb gehabt hätte, wie im Berichtsjahr. Er ist also:

$$P^{P}_{a,b} := \frac{M(b \mid b)}{M(a \mid b)}$$

Dasselbe Spiel kann man natürlich nicht nur mit den Preisen machen, sondern auch für die Mengen. In völliger Analogie definiert man den *Mengenindex nach Laspeyres* bzw. *nach Paasche* durch:

$$Q^L_{a,b} := \frac{M(a \mid b)}{M(a \mid a)}$$
$$Q^P_{a,b} := \frac{M(b \mid b)}{M(b \mid a)}$$

Anhand dieser Definitionen sieht man sofort, dass wir die Entwicklung des Gesamtwertes damit in zwei Bestandteile zerlegt haben: in die Entwicklung der Preise und die Entwicklung der Mengen. Dies drückt sich in folgender Formel aus:

$$P^L_{a,b} Q^P_{a,b} = W_{a,b} = P^P_{a,b} Q^L_{a,b}$$

Wir studieren ein kleines Beispiel: das Sozialprodukt in den Jahren 0 bis 3. Es setzt sich (als Summe) zusammen aus dem Warenwert an produzierten Konsumgütern und Investitionsgütern. Die Preise der Konsum- bzw. Investitionsgüter im Jahr t bezeichnen wir mit $p_{1,t}$ bzw. mit $p_{2,t}$. Und die produzierten Mengen an Konsum- bzw. Investitionsgütern bezeichnen wir analog mit $q_{1,t}$ bzw. mit $q_{2,t}$. Das Sozialprodukt des Jahres t ist also $S_t = p_{1,t}q_{1,t}+p_{2,t}q_{2,t}$. Nehmen wir die folgenden Zahlenwerte an:

t	$p_{1,t}$	$q_{1,t}$	$p_{2,t}$	$q_{2,t}$	S_t
0	2,0	250	1,5	200	800
1	1,9	280	1,7	190	855
2	1,8	320	2,1	160	912
3	1,7	360	2,4	170	1020

Um die verschiedenen Preis und Mengenindices auszuwerten berechnet man am einfachsten die *M-Matrix*, d.h. alle Kombinationen $M(s \mid t)$ wie viel die im Jahr t produzierten Waren im Jahr s wert gewesen wären. Es ergibt sich:

$M(s \mid t)$	$t=0$	$t=1$	$t=2$	$t=3$
$s=0$	800	845	880	975
$s=1$	815	855	880	973
$s=2$	870	903	912	1005
$s=3$	905	932	928	1020

Aus dieser Matrix lassen sich die Mengenindices als Verhältnissen entlang der Zeilen und die Preisindices als Verhältnisse entlang der Spalten ablesen. So ist etwa $P^L_{0,3} = 905/800$ und $Q^L_{0,2} = 975/800$. Die Indices nach Paasche beziehen sich auf die Werte im Berichtsjahr, also $P^P_{0,3} = 1020/975$ und $Q^P_{0,3} = 1020/905$. Dies wird durch folgendes Diagramm illustriert:

$$\begin{array}{ccc} M(a \mid a) & \xrightarrow{Q^L_{a,b}} & M(a \mid b) \\ \downarrow P^L_{a,b} & & \downarrow P^P_{a,b} \\ M(b \mid a) & \xrightarrow{Q^P_{a,b}} & M(b \mid b) \end{array}$$

Die Wertindices berechnen sich als Verhältnisse entlang der Diagonalen, also etwa $W_{0,3} = 1020/800$. Und man kann die obige Produktformel an diesen Zahlen sofort nachvollziehen. Klar ist auch: Die Käufer bevorzugen die billigeren Produkte, passen sich also der Preisentwicklung an. Deswegen gilt bei realen Daten zumeist:

$$P^P_{a,b} \;\leq\; P^L_{a,b} \quad \text{(Laspeyres-Effekt)}$$

Der Laspeyres-Index hat den Vorteil leicht interpretierbar zu sein, durch das veränderte Verhalten der Käufer muss aber der Warenkorb gelegentlich angepasst werden. Der Paasche-Index hingegen orientiert sich stets an den aktuellen Verhältnissen, dafür ändert sich die gesamte Indexreihe in jedem Schritt des Bezugsjahres. Weil also keiner dieser Indices alles kann, verwendet man daher gelegentlich die Indices nach *Fisher*:

$$\begin{aligned} P^F_{a,b} &:= \sqrt{P^L_{a,b}\, P^P_{a,b}} \\ Q^F_{a,b} &:= \sqrt{Q^L_{a,b}\, Q^P_{a,b}} \end{aligned}$$

So richtig clever ist das aber eigentlich auch nicht: diese Indices lassen sich noch schlechter interpretieren, man muss den Warenkorb auf dem Laufenden halten und man muss sie in jedem Schritt neu berechnen. Zum Trost erhält man immerhin:

$$W_{a,b} \;=\; P^F_{a,b}\, Q^F_{a,b}$$

Die Indices nach Laspreyres und Paasche erlauben noch eine andere, interessante Interpretation: wir bezeichnen die relative Preisentwicklung der j-ten Ware mit T_j (und die Umsatzentwicklung mit U_j):

$$\begin{aligned} T_j &:= \frac{p_{j,b}}{p_{j,a}} \\ U_j &:= \frac{p_{j,b} q_{j,b}}{p_{j,a} q_{j,a}} \end{aligned}$$

Den Anteil der j-ten Ware am Gesamtumsatz im Jahr t bezeichnen wir mit

$$g_{j,t} \;:=\; \frac{p_{j,t} q_{j,t}}{\sum_{i=1}^{n} p_{i,t} q_{i,t}}$$

Dann ist der Preisindex nach Laspeyres das mit den Gewichten $g_{j,a}$ des Bezugsjahres gewichtete *arithmetische* Mittel der Teurungsraten T_j und der Preisindex nach Paasche das mit den Gewichten $g_{j,b}$ des Berichtsjahres gewichtete *harmonische* Mittel der Teuerungsraten T_j. Der Wertindex hingegen ist wieder das mit den Gewichten $g_{j,a}$ des Bezugsjahres gewichtete arithmetische Mittel der Umsatzentwicklung U_j

$$\begin{aligned} P^L_{a,b} &= \sum_{j=1}^n g_{j,a}\, T_j \\ \frac{1}{P^P_{a,b}} &= \sum_{j=1}^n g_{j,b} \frac{1}{T_j} \\ W_{a,b} &= \sum_{j=1}^n g_{j,a}\, U_j \end{aligned}$$

Soweit zur Theorie der einzelnen Indices; betrachten wir nun die Indices in der Praxis. Wir müssen klären, wie man Indices aneinander hängt, auf ein anderes Bezugsjahr umrechnet und zur Deflationierung verwendet:

- **Verkettung von Indices:** Nehmen wir an, wir haben drei Zeitpunkte $a \le b \le c$ gegeben. Die Entwicklung von a bis b wird durch den Index $I_{a,b}$ beschrieben und die Entwicklung von b bis c durch den Index $I_{b,c}$. Welcher Index beschreibt dann die Gesamtentwicklung von a bis c? Die Antwort liegt auf der Hand: da Indices als Verhältnisse definiert wurden, muss man einfach das Produkt bilden:

  $$\widetilde{I}_{a,c} \;:=\; I_{a,b}\, I_{b,c}$$

  Besonders einfach ist dies am Wertindex zu sehen. Hier ergibt die Verkettung von $W_{a,b}$ vor $W_{b,c}$ gerade wieder $W_{a,c}$ denn:

  $$\widetilde{W}_{a,c} \;:=\; \frac{M(b \mid b)}{M(a \mid a)} \cdot \frac{M(c \mid c)}{M(b \mid b)} \;=\; \frac{M(c \mid c)}{M(a \mid a)} \;=\; W_{a,c}$$

  Leider ist die Situation bei den Preis- und Mengenindices eine andere. Hier ergibt die Verkettung $P^L_{a,b} P^L_{b,c}$ eben *nicht* $P^L_{a,b}$. Deswegen rechnet man gerne in Einjahresschritten, indem man definiert:

  $$\begin{aligned} \widetilde{P}^P_{a,b} &:= \prod_{t=a+1}^{b} P^P_{t-1,t} \;=\; P^P_{a,a+1} \cdot \ldots \cdot P^P_{b-1,b} \\ \widetilde{Q}^L_{a,b} &:= \prod_{t=a+1}^{b} Q^L_{t-1,t} \;=\; Q^L_{a,a+1} \cdot \ldots \cdot Q^L_{b-1,b} \end{aligned}$$

  Diese Indices lassen sich dann nach Konstruktion einfach aneinander hängen, d.h. es gilt $\widetilde{P}^P_{a,b} \widetilde{P}^P_{b,c} = \widetilde{P}^P_{a,c}$ und genauso $\widetilde{Q}^L_{a,b} \widetilde{Q}^L_{b,c} = \widetilde{Q}^L_{a,c}$. Zudem behält man die schöne Eigenschaft $\widetilde{P}^P_{a,b} \widetilde{Q}^L_{a,b} = W_{a,b}$.

- **Umbasierung von Indices:** Wie bereits erklärt besteht bei den Indices nach Laspeyres die Notwendigkeit den Warenkorb gelegentlich zu aktualisieren. Das statistische Bundesamt etwa tut dies alle 5 Jahre. Wir haben also eine Liste mit Indices $I_{a,t}$ für das Bezugsjahr a und wollen diese zu $\widetilde{I}_{b,t}$ auf das Bezugsjahr b umschreiben. Die Verkettung $I_{a,b}\widetilde{I}_{b,t}$ sollte natürlich wieder $I_{a,t}$ ergeben. Daher definiert man den umbasierten Index durch den Umweg über das alte Bezugsjahr:

  $$\widetilde{I}_{b,t} := \frac{I_{a,t}}{I_{a,b}}$$

  Betrachten wir ein Beispiel: die Preisindices $P^L_{0,t}$ nach Laspeyres für das obige Beispiel des Sozialproduktes. Wir basieren die Indexreihe von Jahr 0 auf Jahr 1 um und vergleichen dies mit dem Wert von $P^L_{1,t}$:

t	$P^L_{0,t}$	$\widetilde{P}^L_{1,t}$	$P^L_{1,t}$
0	1	0,982	0,988
1	1,019	1	1
2	1,088	1,067	1,018
3	1,131	1,110	1,090

- **Deflationierung:** (= Inflationsbereinigung) Die Kaufkraft von Geld ist virtuell, d.h. Geld hat keinen inhärenten Wert, es wird nur aufgrund eines gesellschaftlichen Konsens als werthaltig behandelt. Wenn sich das Verhältnis zwischen Geldmenge und Gesamtwert aller Güter verschiebt, ändert sich also die Kaufkraft eines Euros. Dies äussert sich darin, dass sich der Preis für ein und dieselbe Ware ändert. Eben dies misst aber ein Preisindex. Haben wir im Bezugsjahr a also eine Geldmenge Y_a zur Verfügung, so berechnet sich deren Kaufkraft als $Q = Y_a/P_a$. Der Preis der Waren hat sich bis zum Berichtsjahr b aber nach $P_b = P_aP_{a,b}$ (mit einem passenden Preisindex $P_{a,b}$) verändert. Um dieselben Waren Q zu kaufen benötigt man im Jahr b also die Geldmenge $Y_b = QP_b = QP_aP_{a,b} = Y_aP_{a,b}$. Wollen wir also die Kaufkraft einer Geldmenge Y_b im Jahr b auf das Jahr a zurück rechnen, erfolgt dies nach der Vorschrift:

  $$Y_a = \frac{Y_b}{P_{a,b}}$$

  Dies geschieht bei einer Inflationsbereinigung: man rechnet die gegenwärtige (*nominale*) Geldmenge Y_b im Berichtsjahr b auf das Bezugsjahr a um, und erhält die Geldmenge Y_a, die die (*reale*) Kaufkraft bezogen auf das Jahr a angibt. In der amtlichen Statistik verwendet man als *Deflator* den Preisindex nach Paasche, d.h. man berechnet

  $$Y_a = \frac{Y_b}{\widetilde{P}^P_{a,b}}$$

  Wir führen die Deflationierung einmal am Beispiel des Sozialprodukts aus obigem Beispiel vor: der nominale Wert $S^n_t = M(t \mid t)$ ist der Geldbetrag im jeweiligen Jahr t. Der reale Wert S^r_t ist der mit $\widetilde{P}^P_{0,t}$ auf das Jahr 0 deflationierte Wert:

t	$P^P_{t-1,t}$	$\widetilde{P}^P_{0,t}$	S^n_t	S^r_t
0	n/a	1	800	800
1	1,012	1,012	855	844,9
2	1,036	1,049	912	869,4
3	1,015	1,064	1020	958,6

Beispiel: Hinter der bekannten Inflationsrate verbirgt sich natürlich ein Index. Sie wird ermittelt, indem die Kosten eines fixierten Warenkorbes über die Jahre hinweg berechnet werden (der Warenkorb wird alle 5 Jahre auf den durchschnittlichen Konsum aktualisiert, dann muss der Index umbasiert werden). Es handelt sich dabei also um einen Preisindex nach Laspeyres. Wir geben die Inflationsrate für die vergangenen Jahre an, verketten diese zum Index $\widetilde{P}^L_{2001,t}$ und basieren den um, auf das Jahr 2005

t = Jahr	Inflationsrate	$P^L_{t-1,t}$	$\widetilde{P}^L_{2001,t}$	$\widetilde{P}^L_{2005,t}$
2002	1,5 %	1,015	1,015	0,959
2003	1 %	1,01	1,013	0,969
2004	1,7 %	1,017	1,043	0,985
2005	1,5 %	1,015	1,058	1
2006	1,6 %	1,016	1,075	1,016
2007	2,3 %	1,023	1,020	1,039
2008	3,1 %	1,031	1,134	1,716

Das heißt, hätten Sie einen Betrag von 100 Euro im Jahr 2002 gehabt, so hatte dieser im Jahr 2008 dieselbe Kaufkraft gehabt wie 113,4 Euro. Umgekehrt sind 100 Euro aus dem Jahr 2008 nur soviel wert, wie 88,18 Euro im Jahr 2002.

Kapitel 12

Wahrscheinlichkeiten

Beispiel: Wir werfen 2 normale (6-seitige) Spielwürfel und addieren deren Augensumme. Dann gibt es 36 mögliche Würfelereignisse - die gewürfelten Paare $(1,1)$, $(1,2)$ und so weiter bis $(6,6)$. Jedes dieser Ereignisse ist gleich wahrscheinlich, kommt (im Durchschnitt) also jedes 36te Mal vor. Die möglichen Ergebnisse sind aber die Zahlen von $2 = 1+1$ bis $12 = 6+6$. Die einzige Möglichkeit eine 2 zu erwürfeln ist die Kombination $(1,1)$. Also hat auch die 2 eine Wahrscheinlichkeit von $1/36$. Der Zahl 4 liegen aber 3 mögliche Ereignisse zugrunde - die Paare $(1,3)$, $(2,2)$ und $(3,1)$. Die 4 hat also die Wahrscheinlichkeit $3/36 = 1/12$. Also gerade weil jedes *Elementarereignis* (= das gewürfelte Zahlenpaar) die gleiche Wahrscheinlichkeit hat, haben die *Ergebnisse* (= die Augensummen) verschiedene Wahrscheinlichkeiten.

Definition: Wir bezeichnen die Menge der **Elementarereignisse** mit Ω, die Menge der **Beobachtungswerte** mit S. Eine Funktion $x : \Omega \to S$, die jedem Elementarereignis $t \in \Omega$ ein Ergebnis $x_t \in S$ zuordnet heißt **Zufallsvariable**. Wir sprechen von einem **Laplace-Prozess**, falls

(1) $\Omega = 1 \dots n$ endlich ist, und

(2) alle Elementarereignisse $t \in \Omega$ gleich wahrscheinlich sind.

Im folgenden werden wir nur Laplace-Prozesse betrachten, auch wenn die entwickelte Theorie oft allgemeiner ist. Die Wahrscheinlichkeit eines jeden Elementarereignisses beträgt also $1/n$. Die Wahrscheinlichkeit das Ergebnis $s \in S$ zu beobachten, ergibt sich also zu:

$$p(s) \;:=\; \frac{\text{Zahl der günstigen Fälle}}{\text{Zahl aller möglichen Fälle}} \;=\; \frac{\#\{\, t \in 1 \dots n \mid x_t = s \,\}}{n} \;=\; h_s$$

Bemerkung: Dies ist also bereits die Anknüpfung zur Statistik - die Wahrscheinlichkeit den Wert $s \in S$ zu beobachten ist die relative Häufigkeit des Wertes s unter allen Elementarereignissen. Man kann einen unbekannten Prozess also folgenderma/ssen untersuchen: man macht viele Stichproben, diese liefern die Werte x_1, x_2 bis x_n. Die s_1, s_2 bis s_m seien wieder die darunter vorkommenden, *verschiedenen* Werte. Dann ist $p(s_i) \approx h_i$ statistisch zu bestimmen. Und die Annäherung wird umso besser sein, je mehr Stichproben wir machen, d.h. je größer n wird.

Annahme: Ist $A = \{ s_1, \ldots, s_k \} \subseteq S$ eine Liste möglicher Beobachtungswerte, dann ist die Wahrscheinlichkeit bei *einer* Stichprobe den Wert s_1 *oder* s_2 *oder* ... *oder* s_k zu erhalten offenbar gegeben durch

$$P(A) = \sum_{i=1}^{k} p(s_i)$$

Machen wir hingegen unter immer gleichen Bedingungen k Stichproben *hintereinander* so ist die Wahrscheinlichkeit bei der 1ten Stichprobe den Wert s_1, bei der 2ten Stichprobe den Wert s_2 ... und bei der $k$ten Stichprobe den Wert s_k zu ziehen gegeben, durch $p(s_1) \cdot p(s_2) \cdot \cdots \cdot p(s_k)$.

Beispiel: Wir haben eine Urne mit p weißen und q schwarzen Kugeln vorliegen und es sei $n = p + q$. Die Wahrscheinlichkeit bei 3 Ziehungen, mit Zurücklegen, erst eine weiße, dann eine schwarze und wieder eine weiße Kugel zu ziehen beträgt

$$p(s)p(w)p(s) = \frac{p}{n} \cdot \frac{q}{n} \cdot \frac{p}{n} = \frac{p^2 q}{n^3}$$

Nun dasselbe ohne Zurücklegen - eine gezogene Kugel bleibt draußen. Ziehen wir die erste weiße Kugel, bleiben nur $p - 1$ weiße Kugeln zurück. Die Wahrscheinlichkeit nun eine schwarze Kugel zu ziehen hat sich also geändert, zu $q/(n-1)$. Wurde auch diese Kugel gezogen sind also nur noch $q - 1$ schwarze Kugeln in der Urne. Die Wahrscheinlichkeit, dass die nächste gezogene Kugel weiß ist, ist jetzt also $(p - 1)/(n - 2)$. Insgesamt beträgt die Wahrscheinlichkeit der Zugfolge also

$$\frac{p}{n} \cdot \frac{q}{n-1} \cdot \frac{p-1}{n-2}$$

Axiome der Wahrscheinlichkeiten:

Definition: Will man eine Theorie der Wahrscheinlichkeiten aufbauen, so ist es notwendig einige grundlegende Eigenschaften, wie sich Wahrscheinlichkeiten verhalten, als Annahmen (Axiome) zu Grunde zu legen. Aus diesen Eigenschaften folgert man dann weitere - weniger offensichtliche - Eigenschaften. An den Ausführungen oben erkennt man, dass eine Wahrscheinlichkeit für jede Teilmenge $A \subseteq S$ angegeben werden kann $P(A)$ ist die Wahrscheinlichkeit, dass eines der Ereignisse $s \in A$ eintritt. Formal ist eine **Wahrscheinlichkeitsfunktion** auf S also eine Abbildung der Form $P : \mathcal{P}(S) \to \mathbb{R}$, die die folgenden Eigenschaften erfüllt:

(1) $P(S) = 1$,

(2) für $A \subseteq S$ gilt stets $P(A) \geq 0$ und

(3) für $A, B \subseteq S$ mit $A \cap B = \emptyset$ gilt stets $P(A \cup B) = P(A) + P(B)$

Eigenschaft (1) besagt, dass die Wahrscheinlichkeit bei einer Stichprobe irgendein Ergebnis zu erhalten gleich 1 ist. Eigenschaft (2) besagt, dass es keine negativen Wahrscheinlichkeiten gibt und Eigenschaft (3) ist eine formale Fassung der obigen Annahme.

Satz: Ist nun P eine beliebige Wahrscheinlichkeitsfunktion auf S, so kann man aus den obigen Eigenschaften (1) bis (3) gleich einige weitere Eigenschaften (4) bis (7) für beliebige Teilmengen A, $B \subseteq S$ folgern:

(4) es ist $P(\emptyset) = 0$

(5) es ist $P\left(\overline{A}\right) = 1 - P(A)$

(6) ist $A \subseteq B$ so folgt $P(A) \leq P(B)$

(7) es gilt immer $P(B \setminus A) = P(B) - P(A \cap B)$

(8) es gilt immer $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Definition: Sei nun also $P : \mathcal{P}(S) \to [0,1]$ eine Wahrscheinlichkeitsfunktion auf S. Dann nennen wir eine Abbildung der Form $X : S \to \mathbb{R}$ eine **Zufallsvariable** auf S. Und für diese (und $x \in \mathbb{R}$) bezeichnen wir die Mengen:

$$\begin{aligned} \{\, X = x \,\} &:= \{\, s \in S \mid X(s) = x \,\} \\ \{\, X \leq x \,\} &:= \{\, s \in S \mid X(s) \leq x \,\} \end{aligned}$$

Wir bezeichnen die Wahrscheinlichkeit, dass X den Wert x annimmt mit $P(X = x)$ und die Wahrscheinlichkeit, dass X höchstens den Wert x annimmt, mit $P(X \leq x)$. Formal bedeutet das

$$\begin{aligned} P(X = x) &:= P\left(\{\, s \in S \mid X(s) = x \,\}\right) \\ P(X \leq x) &:= P\left(\{\, s \in S \mid X(s) \leq x \,\}\right) \end{aligned}$$

Kapitel 13

Bedingte Wahrscheinlichkeit

Beispiel: Ihr neuer Geschäftsfreund hat zwei Kinder, eines davon ist ein Mädchen. Wie groß ist die Wahrscheinlichkeit dafür, das das andere Kind ein Junge ist. Naiv würde man sofort 50% sagen. Eine genauere Betrachtung zeigt aber, das das nicht stimmt. Bei 2 Kindern gibt es 4 Möglichkeiten was für ein Geschlecht diese haben: (w,w), (w,m), (m,w) und (m,m). Die Kombination (m,m) können wir im Vorfeld ausschließen, da wir ja bereits wissen, das eines davon ein Mädchen ist. Es bleiben also die 3 Möglichkeiten (w,w), (w,m) und (m,w). In 2 dieser Fälle ist aber ein Junge darunter, nur in einem Fall sind beides Mädchen. Da alle 4 Fälle gleich wahrscheinlich sind, erhält man: die Wahrscheinlichkeit, dass das andere Kind ein Junge ist, ist 2/3!

Beispiel: Kehren wir zurück zu dem Beispiel in Kapitel 8. Wenn wir auswerten wollen, wie hoch die Wahrscheinlichkeit für einen Akademiker ist arbeitslos zu werden, nehmen wir die Zahl der arbeitslosen Akademiker (9) und teilen sie durch die Gesamtzahl aller Akademiker (217), denn das ist ja die Häufigkeit der Arbeitslosen unter den Akademikern. Mit anderen Worten

$$\begin{aligned} P(\text{arbeitslos} \mid \text{Akademiker}) &= \begin{array}{c}\text{Wahrscheinlichkeit für einen}\\ \text{Akademiker arbeitslos zu sein}\end{array} \\ &= \frac{\#\text{arbeitslose Akademiker}}{\#\text{Akademiker}} \\ &= \frac{P(\text{arbeitslos und Akademiker})}{P(\text{Akademiker})} \end{aligned}$$

Definition: Entsprechend dem obigen Vorbild setzen wir: ist P eine Wahrscheinlichkeitsfunktion auf S und sind sind $A, B \subseteq S$ zwei Ereignis(mengen), dann definieren wir die **bedingte Wahrscheinlichkeit**, dass das *Ereignis* A unter der *Bedingung* B eintritt, durch:

$$P(A \mid B) \;:=\; \frac{P(A \cap B)}{P(B)}$$

Und die Ereignisse A und $B \subseteq S$ heißen **(stochastisch) unabhängig**, wenn die Wahrscheinlichkeit von A gar nicht von der Bedingung B abhängt. Formal: die folgenden drei Aussagen (a), (b) und (c) sind äquivalent. Und sind sie erfüllt (es genügt wenn eine dies ist), dann nennen wir A und B (stochastisch) unabhängig:

(a) $P(A \cap B) \;=\; P(A) \cdot P(B)$

(b) $P(A \mid B) \;=\; P(A)$

(c) $P(B \mid A) \;=\; P(B)$

Satz von der totalen Wahrscheinlichkeit: Es sei P eine Wahrscheinlichkeitsfunktion auf S, $A \subseteq S$ ein beliebiges Ereignis und S_1, S_2 bis S_m sei eine Klassierung von S (siehe Kapitel 7). Dann lässt sich die Wahrscheinlichkeit von A berechnen, nach:

$$P(A) \;=\; \sum_{i=1}^{m} P(A \mid S_i)\, P(S_i)$$

Beispiel: Bei einer Spielshow wählt der Kandidat eines von 3 Toren A, B oder C. Hinter einem der Tore ist der Gewinn, zwei der Tore sind Nieten. Die Chance auf das richtige Tor zu tippen ist also 1/3. Nachdem der Kandidat gewählt hat, öffnet der Showmaster ein drittes Tor - weder das gewählte, noch das mit dem Gewinn - und fragt 'Wollen Sie jetzt doch das andere Tor nehmen?' Was würden Sie tun? Denken wir nach: wenn Sie zuerst richtig gewählt hatten (= Bedingung *richtig*), dann müssen Sie das richtige Tor jetzt verlassen. Ihre Gewinnchance (= Ereignis *Gewinn*) ist in diesem Fall also gleich 0:

$$P(\text{Gewinn} \mid \text{richtig}) \;=\; 0$$

Hatten Sie aber zuerst das falsche Tor gewählt (= Bedingung *falsch*), dann weichen Sie jetzt zwingend auf das richtige Tor aus (denn eine Niete verlassen Sie und eine Niete hat Ihnen der Showmaster gezeigt, es bleibt nur das Gewinntor). Ihre Gewinnchance ist in diesem Fall also gleich 1:

$$P(\text{Gewinn} \mid \text{falsch}) \;=\; 1$$

Die beiden Fälle *richtig* und *falsch* bilden offensichtlich eine Klassierung des Raumes S. Es gilt also der Satz der totalen Wahrscheinlichkeit:

$$\begin{aligned} P(\text{Gewinn}) \;&=\; P(\text{Gewinn} \mid \text{richtig})P(\text{richtig}) \\ &\quad +P(\text{Gewinn} \mid \text{falsch})P(\text{falsch}) \\ &=\; 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} \;=\; \frac{2}{3} \end{aligned}$$

Wenn Sie sich umentscheiden steigt Ihre Gewinnchance also auf 2/3! Allgemeiner kann man diese Spiel mit n Toren spielen. Dann sind die bedingten Wahrscheinlichkeiten wieder $P(\text{Gewinn} \mid \text{richtig}) = 0$ und (weil 2 Nieten entfallen) $P(\text{Gewinn} \mid \text{falsch}) = 1/(n-2)$. Insgesamt also:

$$P(\text{Gewinn}) \;=\; 0 \cdot \frac{1}{n} + \frac{1}{n-2} \cdot \frac{n-1}{n} \;=\; \frac{n-1}{n(n-2)}$$

Satz von Bayes: Sei wieder P eine Wahrscheinlichkeitsfunktion auf S und $A \subseteq S$ ein beliebiges Ereignis und S_1, S_2 bis S_m sei eine Klassierung von S. Dann gilt umgekehrt auch:

$$P(S_k \mid A) \;=\; \frac{P(A \mid S_k)P(S_k)}{P(A)} \;=\; \frac{P(A \mid S_k)P(S_k)}{\sum_{i=1}^{m} P(A \mid S_i)\, P(S_i)}$$

Beispiel: Jetzt wird es leicht politisch: wir untersuchen die Beweiskraft von Geständnissen. Wir bezeichnen die die folgenden Ereignisse: $S :=$ der Angeklagte ist schuldig, $U :=$ der Angeklagte ist unschuldig und $G :=$ der Angeklagte hat ein Geständnis abgelegt. Offensichtlich bilden S und U eine Klassierung des Raumes und damit können wir die Wahrscheinlichkeit, das der Angeklagte schuldig ist, unter der Annahme das er gestanden hat, mit dem Satz von Bayes ausdrücken:

$$P(S \mid G) = \frac{P(G \mid S)P(S)}{P(G \mid U)P(U) + P(G \mid S)P(S)}$$

Es bezeichne weiterhin $p := P(S)$ und $r := P(G \mid U)/P(G \mid S)$. D.h. p ist die (unbekannte) Schuldwahrscheinlichkeit des Angeklagten und r gibt an um wieviel (un)wahrscheinlicher es ist, dass ein Unschuldiger gesteht, als ein Schuldiger. Wegen $P(U) = 1 - p$ erhalten wir (durch Kürzen von $P(G \mid S)$):

$$P(S \mid G) = \frac{p}{r(1-p) + p}$$

Das Gericht ist natürlich davon überzeugt, dass ein Geständnis dafür spricht, dass der Angeklagte tatsächlich schuldig ist. D.h. der Richter glaubt an $P(S) \leq P(S \mid G)$. Eine leichte Rechnung zeigt (mit obiger Formel), dass dies gerade $r \leq 1$, also $P(G \mid U) \leq P(G \mid S)$ bedeutet. Das heißt ein Geständnis erhöht die Schuldwahrscheinlichkeit nur, wenn eine schuldige Person eher gesteht, als eine unschuldige Person. Auf den ersten Blick denkt man: ja das ist doch wohl so. Aber stimmt das überhaupt? Es kommt tatsächlich vor, dass ein Unschuldiger unter Druck Taten gesteht, die er gar nicht begangen hat. Und wenn wir an Terroristen denken - die wurden in ihren Trainigscamps ausgebildet Druck auszuhalten, der Normalbürger nicht. Mit einem Blick auf Guantanamo sollte man also fest halten: bei mutmaßlichen Terroristen *senkt* ein Geständnis, das unter Druck gegeben wurde, die Schuldwahrscheinlichkeit!

Kapitel 14

Kombinatorik

In der Kombinatorik fragt man sich immer, wieviele Möglichkeiten es gibt eine bestimmte Situation zu realisieren. Die Frage, wie viele verschiedene Elementarereignisse es gibt, ist also immer eine kombinatorische Frage. Wir betrachten der Reihe nach verschiedene Grundprobleme:

1. **Schalterstellungen:** Wir haben n Schalter vorliegen. Jeder dieser Schalter besitzt k mögliche Stellungen. Wieviele Schaltereistellungen sind dann möglich? Die Antwort liegt auf der Hand: n^k viele. Dies Problem kann auch so formuliert werden: wie viele Möglichkeiten gibt es Worte mit k Buchstaben Länge zu bilden, wenn das Alphabet n verschiedene Zeichen kennt? Wieder n^k. Noch eine Formulierung desselben Problems: wir haben eine Urne mit n nummerierten Kugeln. Aus dieser ziehen wir k Mal, notieren die Nummer und legen die Kugel zurück. Wie viele verschiedene Ziehungen gibt es? Wieder n^k.

2. **Lottoziehung:** Wir haben n nummerierte Kugeln in einer Urne. Aus dieser ziehen wir k Mal und legen die Kugeln nicht wieder zurück (sondern in der Reihenfolge der Ziehung vor uns hin). Wie viele mögliche Ziehungen gibt es?

$$n(n-1)(n-2)\dots(n-k+1) \quad = \quad \frac{n!}{(n-k)!} \quad = \quad k!\binom{n}{k}$$

   Ein wichtiger Spezialfall ist der Folgende: wollen wir die n Kugeln umordnen, so können wir dies tun, indem wir Eine nach der Anderen ziehen, bis alle n Stück gezogen wurden. D.h. es ist $k = n$ und damit gibt es $n!$ mögliche Anordnungen (**Permutationen**) der n Kugeln.

3. **Lottoauswertung:** Beim Lotto werden die Kugeln danach entsprechend ihrer Größe sortiert, d.h. die Reihenfolge der Ziehung geht verloren. Je $k!$ verschiedene Lottoziehungen ergeben also ein und dasselbe Lottoergebnis. Entsprechend finden wir:

$$\binom{n}{k} \quad = \quad \frac{n}{1}\cdot\frac{n-1}{2}\cdot\dots\cdot\frac{n-k+1}{k}$$

Dieses Problem lässt sich auch so formulieren: wir haben n nummerierte Kugel in Urne 1. Aus dieser Urne 1 ziehen wir k Mal (ohne Zurücklegen) und legen die gezogene Kugel in eine zweite Urne 2. Dann gibt es wieder $\binom{n}{k}$ Möglichkeiten Urne 2 zu füllen.

4. **Zweikugelordnung:** Wir haben N Kugeln gegeben, n weiße und $m = N - n$ schwarze. Wieviele Arten gibt es diese Kugeln anzuordnen (d.h. in eine Reihe zu legen)? Die Antwort lautet:

$$\binom{N}{n} = \binom{n+m}{n}$$

Begründung: wir schreiben die Nummern 1 bis N auf die Kugeln. Es gibt $N!$ Möglichkeiten die nummerierten Kugeln anzuordnen. Wischen wir die Nummern jetzt wieder weg, so spielt es aber keine Rolle mehr, in welcher Reihenfolge die weißen (bzw. die schwarzen) Kugeln untereinander liegen. Je $n!$ für die Weißen und $(N-n)!$ für die Schwarzen Anordnungen sind also identisch. D.h. die $N!$ Anordnungen der nummerierten Kugeln reduzieren sich auf $N!/(n!(N-n)!) = \binom{N}{n}$ Anordnungen für die unnummerierten Kugeln.

5. **Kugelverteilung I:** Wir haben n nummerierte Kugeln und k Urnen vorliegen. Wir wollen nun die Kugeln so auf die Urnen verteilen, das in die erste Urne n_1 Kugeln kommen, und ... und in die $k$te Urne n_k Kugeln kommen ($n = n_1 + n_2 + \cdots + n_k$). Wieviele Möglichkeiten gibt es das zu tun?

$$\frac{n!}{n_1!\, n_2! \cdot \ldots \cdot n_k!}$$

Begründung: es gibt $n!$ Möglichkeiten die n Kugeln anzuordnen. Die ersten n_1 kommen in Urne 1, ..., die letzten n_k kommen in Urne k. Die Anordnung innerhalb der Urnen ist aber egal. D.h. wir müssen durch $n_1!$ (für Urne 1), und ... und durch $n_k!$ (für Urne k) dividieren.

6. **Kugelverteilung II:** Wir haben n identische Kugeln und k Urnen gegeben. Wir wollen diese Kugeln irdendwie auf die Urnen verteilen. Wieviele verschiedene Verteilungen gibt es?

$$\binom{n+k-1}{n}$$

Begründung: Wir führen eine günstige Art ein, die Verteilungen zu codieren: eine weisse Kugel $\circ$ codiert eine der n Kugel, und eine schwarze Kugel $\bullet$ codiert eine Trennwand. Zum Beispiel kodiert $\circ \circ \bullet \bullet \circ$ folgende Situation: es gibt 3 Kugeln und 3 Urnen. Die erste Urne enthält 2 Kugeln, die zweite Urne ist leer und die dritte Urne enthält eine Kugel. Wir haben also n weisse Kugeln und (bei k Urnen) $m = k-1$ schwarze Kugeln zur Codierung. Nach (4) gibt es genau $\binom{n+m}{n}$ solche Codes, also $\binom{n+k-1}{n}$ Verteilungen.

7. **Kugelziehen II:** Wir haben eine Urne mit n nummerierten Kugeln. Aus dieser ziehen wir k Mal (mit Zurücklegen) und notieren die Häufigkeiten, wie oft welche Nummer gezogen wurde. Wieviele Häufigkeitsverteilungen gibt es?

$$\binom{n+k-1}{k}$$

Begründung: Zu jeder Kugelnummer nehmen wir eine Urne. Und wir stellen einen Sack mit k gleichartigen Bällen bereit. Wir ziehen nun eine Kugel aus der Urne, legen sie wieder zurück und tun dafür einen Ball aus dem Sack in die Urne der gezogenen Kugel. Das Ganze machen wir k Mal. Die Zahl der Bälle in den Urnen gibt also die Häufigkeitsverteilung an. Dazu haben wir aber k identische Bälle auf n Urnen verteilt. Und nach (6) gibt es dabei $\binom{k+n-1}{k}$ Möglichkeiten.

Kapitel 15

Zusammenfassung

Übersicht:

	qualitativ	komparativ	quantitativ
Lagemaß	Modus $D(x)$	Median $Z(x)$	Durchschnitt $A(x)$
Streuungsmaß	Dispersion P	Diverstät D	Q, $\sigma_1(x)$ und $\sigma_2(x)$
Konzentration	n/a	n/a	Gini-Koeffizient R
Zusammenhang	Kontingenz C^*	Rangkorrelation	Korrelation r

Lagemaße:
Wir betrachten n *Merkmalsträger*, nummeriert mit $t \in 1 \ldots n$. Die *Ausprägung* des t-ten Merkmalsträgers wird mit x_t bezeichnet. Die möglichen, verschiedenen Ausprägung seien s_i wobei $i \in 1 \ldots m$. Die *absolute Häufigkeit* der Ausprägung s_i wird mit n_i, die *relative Häufigkeit* mit $h_i = n_i/n$ bezeichnet. Dann gilt:

$$\begin{aligned}
X &:= \sum_{t=1}^{n} x_t = \sum_{i=1}^{m} n_i s_i \\
A(x) &:= \frac{1}{n}\sum_{t=1}^{n} x_t = \sum_{i=1}^{m} h_i s_i \\
D(x) &:= s_k \quad \text{so dass} \quad n_k = \max\{\, n_i \mid i \in 1 \ldots m \,\} \\
Z(x) &:= x_k \quad \text{wobei} \quad k = \begin{cases} \frac{n}{2} & \text{für } n \text{ gerade} \\ \frac{n+1}{2} & \text{für } n \text{ ungerade} \end{cases}
\end{aligned}$$

Konzentration:
Im Fall eines komparativen Merkmals sei $s_1 \leq s_2 \leq \cdots \leq s_m$. Dann bezeichnen N_k bzw. H_k die *kumulierten absoluten* bzw. *relativen Häufigkeiten* und L_k die *kumulierte, relative Merkmalssumme* (wobei $k \in 0 \ldots m$)

$$\begin{aligned}
N_k &:= \sum_{i=1}^{k} n_i = N_{k-1} + n_k \\
H_k &:= \frac{N_k}{n} = \sum_{i=1}^{k} h_i = H_{k-1} + h_k
\end{aligned}$$

$$
\begin{aligned}
\ell_i &:= \frac{n_i s_i}{X} &= \frac{h_i s_i}{A(x)} \\
L_k &:= \sum_{i=1}^{k} \ell_i &= L_{k-1} + \ell_k
\end{aligned}
$$

Der *Gini-Koeffizient* R ist dann definiert als das Doppelte der Fläche zwischen Lorenz-Kurve der Verteilung und der Diagonalen. Es gilt

$$
R = 1 - \sum_{i=1}^{m} h_i(L_{i-1} + L_i) = \frac{2}{nX} \sum_{t=1}^{n} t x_t - \frac{n+1}{n}
$$

$$
\underbrace{0 \bullet \bullet 0.25}_{\text{gute Gleichverteilung}} \quad \bullet \bullet \bullet \quad \underbrace{0.4 \bullet \bullet \bullet \bullet 1}_{\text{starke Ungleichverteilung}}
$$

Streuungsmaße:
Im Falle eines quantitativen Merkmals misst man die Streuung der Verteilung mit der *mittleren absoluten Abweichung* $\sigma_1(x)$ oder der *mittleren quadratischen Abweichung (= Standardabweichung)* $\sigma_2(x)$:

$$
\begin{aligned}
\sigma_1(x) &:= \frac{1}{n} \sum_{t=1}^{n} |x_t - Z(x)| \\
\sigma_2(x) &:= \sqrt{\frac{1}{n} \sum_{t=1}^{n} (x_t - A(x))^2} \\
\sigma_2^2(x) &:= \frac{1}{n} \sum_{t=1}^{n} (x_t - A(x))^2 \\
&= \sigma_2(x)^2 = A\big(x^2\big) - A(x)^2
\end{aligned}
$$

Bei qualitativen Merkmalen muss man sich auf die Häufigkeiten zurück ziehen, es bleiben die *Dispersion* P und die *Diversität* D

$$
\begin{aligned}
P &:= \frac{m}{m-1} \left(1 - \sum_{i=1}^{m} h_i^2 \right) \\
D &:= \frac{4}{m-1} \sum_{i=1}^{m} H_i (1 - H_i)
\end{aligned}
$$

$$
\underbrace{0 \bullet \bullet \bullet \bullet \bullet \bullet 0.8}_{\text{starke Ballung}} \bullet \bullet \quad \underbrace{0.9 \bullet 1}_{\text{starke Streuung}}
$$

$$
\underbrace{0 \bullet \bullet \bullet \bullet 0.6}_{\text{starke Ballung}} \bullet \bullet \bullet \quad \underbrace{0.8 \bullet \bullet 1}_{\text{starke Streuung}}
$$

Klassierte Verteilungen:
Im Fall einer klassierten Verteilung mit den Klassen $S_i = [a_{i-1}\,, a_i\,[$ (mit $i \in 1 \ldots m$) bezeichnen wir die *absolute* bzw. *relative Häufigkeit* der Klasse S_i wieder mit n_i bzw. h_i. Ferner bezeichnet man die *Klassenweiten* w_i *Häufigkeitsdichte* h_i^* und *Klassenmitten* s_i^*

$$
\begin{aligned}
w_i &:= a_i - a_{i-1} \\
h_i^* &:= \frac{h_i}{w_i} \\
s_i^* &:= \frac{a_{i-1} + a_i}{2} \\
X_i^* &:= n_i \int_{a_{i-1}}^{a_i} h_i^* s \, ds \;=\; n_i s_i^* \\
X^* &:= \sum_{i=1}^{m} X_i^* \;=\; \sum_{i=1}^{m} n_i s_i^* \\
\ell_i^* &:= \frac{X_i^*}{X^*} \;=\; \frac{h_i s_i^*}{A^*(x)} \\
L_k^* &:= \sum_{i=1}^{k} \ell_i^* \;=\; L_{k-1}^* + \ell_k^*
\end{aligned}
$$

Damit ergeben sich die folgenden Lagemaße für klassierte Verteilungen (dabei bezeichnet H die *empirische Verteilungsfunktion*, $Z^*(x)$ ist dann der Wert $\overline{x}$, für den $H(\overline{x}) = 1/2$ ist, also $H_{k-1} \leq 1/2 < H_k$)

$$
\begin{aligned}
A^*(x) &:= \frac{1}{n} X^* \;=\; \sum_{i=1}^{m} h_i s_i^* \\
D^*(x) &:= s_k^* \quad \text{so dass} \quad h_k^* = \max\{\, h_i^* \mid i \in 1 \ldots m \,\} \\
Z^*(x) &= a_k - \frac{H_k - 1/2}{h_k^*}
\end{aligned}
$$

Die mittlere (absolute bzw. quadratische) Abweichung und der Gini-Koeffizient bestehen dann aus zwei Anteilen - einem *externen* Anteil (1. Summand), der durch die Verteilung auf die Klassen entsteht und einem *internen* Anteil (2. Summand), der innerhalb der Klassen entsteht:

$$
\begin{aligned}
\sigma_2^*(x)^2 &= \sum_{i=1}^{m} h_i (s_i^* - A^*(x))^2 + \frac{1}{12} \sum_{i=1}^{m} h_i w_i^2 \\
\sigma_1^*(x) &= \sum_{i=1}^{m} h_i |s_i^* - Z^*(x)| + h_k^* \left(\frac{w_k}{2} - |s_k^* - Z^*(x)| \right)^2 \\
&\quad \text{wobei } k \in 1 \ldots m \text{ so, dass } a_{k-1} \leq Z^*(x) \leq a_k \\
R^* &= 1 - \sum_{i=1}^{m} h_i (L_{i-1}^* + L_i^*) + \frac{1}{6A^*(x)} \sum_{i=1}^{m} h_i^2 w_i
\end{aligned}
$$

Kontingenz:
Wir betrachten wieder n Merkmalsträger, nummeriert mit $t \in 1 \ldots n$. Diesmal werden jedem t jedoch 2 Ausprägungen $x_t \in R = \{ r_1, \ldots, r_p \}$ und $y_t \in S = \{ s_1, \ldots, s_q \}$ zugeordnet. Die *absolute* (bzw. *relative*) *Häufigkeit* der Merkmalskombination (r_i, s_j) bezeichnen wir mit $n_{i,j}$ (bzw. $h_{i,j}$). Dann bezeichnen wir die *Randhäufigkeiten* bzw. *bedingten Häufigkeiten*

$$
\begin{aligned}
n_{i,+} &:= \#\{ t \in 1 \ldots n \mid x_t = r_i \} = \sum_{j=1}^{q} n_{i,j} \\
n_{+,j} &:= \#\{ t \in 1 \ldots n \mid y_t = s_j \} = \sum_{i=1}^{p} n_{i,j} \\
h_{i,+} &:= \frac{n_{i,+}}{n} = \sum_{j=1}^{q} h_{i,j} \\
h_{+,j} &:= \frac{n_{+,j}}{n} = \sum_{i=1}^{p} h_{i,j} \\
u_{i,j} &:= h_{+,j} h_{i,+} \\
h(i \mid j) &:= \frac{n_{i,j}}{n_{+,j}} = \frac{h_{i,j}}{h_{+,j}} \\
h(j \mid i) &:= \frac{n_{i,j}}{n_{i,+}} = \frac{h_{i,j}}{h_{i,+}}
\end{aligned}
$$

Die Merkmale x und y sind *unabhängig* wenn für alle $i \in 1 \ldots p$ und alle $j \in 1 \ldots q$ gilt $h_{i,j} = u_{i,j}$. Allgemein misst man den Zusammenhang zwischen den beiden Merkmalen mit der *(korrigierten, quadratischen) Kontingenz* C^* (wobei $C \in 0 \ldots (\mu - 1)$ und $\mu := \min\{ p, q \}$)

$$
\begin{aligned}
C &:= \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(h_{i,j} - u_{i,j})^2}{u_{i,j}} \\
C^* &:= \sqrt{\frac{\mu}{\mu - 1}} \sqrt{\frac{C}{C+1}}
\end{aligned}
$$

Lineare Regression:
Der absolute, lineare Zusammenhand der Merkmale x und y wir mit Hilfe der *Kovarianz* $\sigma(x \mid y)$ gemessen. Normiert auf Standardabweichungen $\sigma(x) = \sigma_2(x)$, bzw. $\sigma(y) = \sigma_2(y)$ wird dies *Korrelationskoeffizient* r genannt

$$
\begin{aligned}
\sigma(x \mid y) &:= \frac{1}{n} \sum_{t=1}^{n} (x_t - A(x))(y_t - A(y)) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{q} h_{i,j} (r_i - A(x))(s_j - A(y))) \\
&= A(xy) - A(x)A(y) \\
r &:= \frac{1}{n} \sum_{t=1}^{n} \frac{x_t - A(x)}{\sigma(x)} \cdot \frac{y_t - A(y)}{\sigma(y)} \\
&= \frac{\sigma(x \mid y)}{\sigma(x)\sigma(y)}
\end{aligned}
$$

$$\underbrace{-1 \;\bullet\; \bullet \; -0.6}_{\text{stark gegenläufiger Zusammenhang}} \quad \bullet \; \bullet \quad \underbrace{-0.4 \;\bullet\; \bullet \; 0.4}_{\text{kaum linearer Zusammenhang}} \quad \bullet \; \bullet \quad \underbrace{0.6 \;\bullet\; \bullet \; 1}_{\text{stark gleichläufiger Zusammenhang}}$$

Die *Regressionsgerade* $y = ax + b$ ist die Gerade, die den Zusammenhang am besten wieder gibt (sie minimiert die Streuung $\sigma(u)^2$ der Abweichung $u_t = y_t - (ax_t + b)$). Sie berechnet sich als

$$\begin{aligned} a &= \tfrac{A(xy)-A(x)A(y)}{A(x^2)-A(x)^2} &&= \frac{\sigma(x \mid y)}{\sigma(x)^2} \\ b &= \tfrac{A(y)A(x^2)-A(x)A(xy)}{A(x^2)-A(x)^2} &&= A(y) - aA(x) \end{aligned}$$

Die Streuung der y zerfällt in $\sigma(y)^2 = \sigma(ax+b)^2 + \sigma(u)^2$. Der Anteil der Streuung der y, der durch die Streuung der x und den linearen Zusammenhang erklärt werden kann ist nun gegeben, durch

$$r^2 = \frac{\sigma(ax+b)^2}{\sigma(y)^2}$$

Zeitreihenanalyse:
Wir betrachten ein Merkmal y im Verlaufe der Zeit $t = (1, 2, \ldots, n)$, wobei $n = km$ in m Zyklen von je k Saisons zerfällt. Die *Trendgerade* $y = at + b$ ist die Regressionsgerade zwischen t und y, es gilt

$$\begin{aligned} A(t) &= \frac{n+1}{2} \\ \sigma(t) &= \sqrt{\frac{n^2-1}{12}} \\ a &= \frac{12}{n^2-1}A(ty) - \frac{6}{n-1}A(y) \\ b &= A(y) - \frac{n+1}{2}a \end{aligned}$$

Wir wollen nun den saisonalen Einfluss s_i schätzen $y_t = at + b + s_i$ (wobei $t = jk + i$). Die Streuung $\sigma(u)^2$ der Abweichung $u_t = y_t - (at + b + s_i)$ wird minimal für die saisonalen Einflüsse

$$s_i = \frac{1}{m}\sum_{j=0}^{m-1}(y_{i+jk} - a(i+jk) - b) = \frac{1}{m}\sum_{j=0}^{m-1} y_{i+jk} - \left(ai + b + \frac{n-k}{2}a\right)$$

Die Streuung der y zerfällt in $\sigma(y)^2 = \sigma(at+b)^2 + \sigma(s)^2 + \sigma(u)^2$. Der Anteil der durch den Trend und die saisonalen Einflüsse erklärt werden kann ist

$$\begin{aligned} r^2 &= \frac{\sigma(at+b)^2 + \sigma(s)^2}{\sigma(y)^2} \\ \sigma(at+b)^2 &= \frac{n^2-1}{12}a^2 \\ \sigma(s)^2 &= \frac{1}{m}\sum_{i=1}^{k} s_i^2 \end{aligned}$$

Rangkorrelation:
Im Fall zweier komparativer Merkmale komparativen Merkmale $x_1 \leq x_2 \leq \cdots \leq x_n$ und $y_1 \leq y_2 \leq \cdots \leq y_n$ vergibt man *Ränge* zur Berechnung der Kovarianz. Der Rang der i-ten Ausprägung berechnen sich nach

$$R(s_i) = N_{i-1} + \frac{n_i + 1}{2}$$

Und damit ergibt sich dann das arithmetische Mittel, bzw. die Kovarianz zu (es macht keinen Sinn eine Korrelation berechnen zu wollen)

$$\begin{aligned} A(R(x)) &= \frac{n+1}{2} \\ \sigma(R(x) \mid R(y)) &= A\big(R(x)R(y)\big) - \left(\frac{n+1}{2}\right)^2 \end{aligned}$$

Kapitel 16

Mathematische Ergänzungen

In diesem Kapitel werden wir manche Aussagen aus den vorangegangenen Kapiteln mathematisch sauber fassen und vor allem alle Formeln beweisen. Dieses Kapitel wendet sich also primär an die mathematisch interessierten Leser, die sich nicht mit einem *das ist halt so* zufrieden geben. Zunächst beweisen wir die Formeln der speziellen Summen:

Satz 16.1: Es seien $1 \leq n \in \mathbb{N}$ und $q \in \mathbb{R}$ mit $q \neq 1$. Dann gelten die folgenden Formeln zur Berechnung einiger einfacher Summen:

$$\begin{aligned}
\sum_{k=1}^{n} k &= \frac{n(n+1)}{2} \\
\sum_{k=1}^{n} (2k-1) &= n^2 \\
\sum_{k=0}^{n} q^k &= \frac{q^{n+1}-1}{q-1} \\
\sum_{k=1}^{n} k^2 &= \frac{(2n+1)(n+1)n}{6}
\end{aligned}$$

Beweis: Die erste Formel beweist man mit einem fast schon zauberhaften Trick: wir schreiben die Zahlen 1, 2 und so weiter bis n in aufsteigender Reihenfolge hin. Darunter schreiben wir sie ein zweites Mal, aber in absteigernder Reihenfolge, also n, $n-1$ bis 1:

$$\begin{array}{cccccc}
1 & 2 & 3 & \cdots & n-1 & n \\
n & n-1 & n-2 & \cdots & 2 & 1
\end{array}$$

Addiert man je zwei übereinander stehende Zahlen, so ergibt sich immer die Summe $n+1$. Insgesamt hat man also die Summe n mal $n+1$. Wir haben die Zahlen 1 bis n aber zwei Mal hingeschrieben, also $2(1+2+\cdots+n) = n(n+1)$. Die Behauptung ergibt sich durch Division durch 2. Die zweite Formel erhl̈t man aus der ersten durch eine einfache Rechnung:

$$\sum_{k=1}^{n} (2k-1) = 2\sum_{k=1}^{n} k - \sum_{k=1}^{n} 1 = 2\frac{n(n+1)}{2} - n = n^2 + n - n = n^2$$

Die dritte Formel erhält man wieder durch einen schönen algebraischen Trick - wir beginnen mit dem folgenden Ausdruck:

$$
\begin{aligned}
(q-1)\sum_{k=0}^{n} q^k &= q\sum_{k=0}^{n} q^k - \sum_{k=0}^{n} q^k = \sum_{k=0}^{n} q^{k+1} - \sum_{k=0}^{n} q^k \\
&= \sum_{k=1}^{n+1} q^k - \sum_{k=0}^{n} q^k = q^{n+1} - q^0 = q^{n+1} - 1
\end{aligned}
$$

Ist $q \neq 1$ kann man diese Gleichung durch $q-1$ dividieren, sofort die Behauptung liefert. Der Beweis der vierten Formel verwendet die Beweismethode der *vollständigen Induktion*. Die Idee dabei ist folgende: zunächst rechnen wir die Aussage für $n = 1$ nach: $(2+1)(1+1)1/6 = 6/6 = 1 = 1^2$ stimmt. Der erste Schritt $n = 1$ ist damit erledigt. Nun setzen wir einen Schritt nach dem anderen. Wenn wir allgemein die Schritte 1, 2 und so weiter bis n schon alle nachgerechnet haben, dann ist $n + 1$ als nächster Schritt dran. Um diesen auch noch nachzurechnen fangen wir an mit:

$$
\begin{aligned}
\Big(2(n+1)+1\Big)\Big((n+1)+1\Big)(n+1) &= (n+1)\Big((2n+3)(n+2)\Big) \\
&= (n+1)\Big(2n^2+7n+6\Big) \\
&= (n+1)\Big((2n+1)n+6(n+1)\Big) \\
&= (2n+1)(n+1)n+6(n+1)^2
\end{aligned}
$$

Wir bezeichnen die Summe $s_n = 1^2 + 2^2 + \cdots + n^2$. Da wir gerade im Schritt $n + 1$ sind, haben wir $s_n = (2n + 1)(n + 1)n/6$ schon nachgerechnet. Und setzen wir die obige Rechnung ein so kommen wir wie gewünscht einen Schritt weiter (und damit gilt die Behauptung für alle n):

$$
\begin{aligned}
s_{n+1} &= s_n + (n+1)^2 = \frac{(2n+1)(n+1)n}{6} + \frac{6(n+1)^2}{6} \\
&= \frac{\Big(2(n+1)+1\Big)\Big((n+1)+1\Big)(n+1)}{6}
\end{aligned}
$$

□

Definition 16.2: Seien $x = (x_1, x_2, \ldots, x_n)$ und $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ zwei n-Tupel reeller Zahlen. Dann definieren wir das **arithmetische Mittel** $A(x)$, die **Varianz** $\sigma^2(x)$ und die **Kovarianz** $\sigma(x \mid y)$ durch:

$$
\begin{aligned}
A(x) &:= \frac{1}{n}\sum_{t=1}^{n} x_t \\
\sigma(x \mid y) &:= \frac{1}{n}\sum_{t=1}^{n}(x_t - A(x))(y_t - A(y)) \\
\sigma^2(x) &:= \sigma(x \mid x)
\end{aligned}
$$

Bemerkung 16.3: Wie zuvor bezeichnen wir die *verschiedenen* Zahlen unter den x_t mit r_i, d.h. $\{ x_1, x_2, \ldots, x_n \} = \{ r_1, r_2, \ldots, r_p \}$ wobei die r_i paarweise verschieden sein sollen. Die *absolute Häufigkeit* n_i von r_i ist die Anzahl der Vorkommen von r_i unter den x_t, formal: $n_i := \# \{ t \in 1 \ldots n \mid x_t = r_i \}$. Die *relative Häufigkeit* von r_i ist $h_i := n_i/n$. Damit lässt sich das arithmetische Mittel berechnen, als:

$$A(x) \;=\; \frac{1}{n}\sum_{t=1}^{n} x_t \;=\; \frac{1}{n}\sum_{i=1}^{p} n_i s_i \;=\; \sum_{i=1}^{p} h_i s_i$$

Genauso bezeichnen wir die *verschiedenen* Zahlen unter den y_t mit s_j. Die möglichen verschiedenen *Paare*, die unter den (x_t, y_t) vorkommen, sind also irgend welche (r_i, s_j). Wir bezeichnen mit $n_{i,j}$ die Zahl der (x_t, y_t), die gleich (r_i, s_j) sind, formal $n_{i,j} := \# \{ t \in 1 \ldots n \mid x_t = r_i \text{ und } y_t = s_j \}$. Damit sieht man dann

$$A(xy) \;=\; \frac{1}{n}\sum_{t=1}^{n} x_t y_t \;=\; \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{i,j} r_i s_j \;=\; \sum_{i=1}^{p}\sum_{j=1}^{q} h_{i,j} r_i s_j$$

Fassen wir die Zahlen $x_t - A(x)$ zum Vektor $x - A(x)$ zusammen (genauso $y - A(y)$), so ist die Kovarianz (nach Definition) nichts anderes als das arithmetische Mittel des Produktvektors $(x - A(x))(y - A(y))$, sprich:

$$\sigma(x \mid y) \;=\; A\Big((x - A(x))(y - A(y))\Big)$$

Satz 16.4: Seien x, x' und $y \in \mathbb{R}^n$ Vektoren und sei $a \in \mathbb{R}$ eine Zahl. Wie üblich interpretieren wir $a = (a, a, \ldots, a) \in \mathbb{R}^n$ auch als Vektor. Dann erfüllt das aritmetische Mittel die folgenden Eigenschaften (genannt *Linearität*):

$$\begin{aligned} A(a) &= a \\ A(ax) &= aA(x) \\ A(x + y) &= A(x) + A(y) \end{aligned}$$

Und die Kovarianz erfüllt ebenfalls einige Eigenschaften, die als *Symmetrie* bzw. *Bilinearität* bezeichnet werden:

$$\begin{aligned} \sigma(a \mid x) &= 0 \\ \sigma(x \mid y) &= \sigma(y \mid x) \\ \sigma(ax \mid y) &= a\sigma(x \mid y) \\ \sigma(x + x' \mid y) &= \sigma(x \mid y) + \sigma(x' \mid y) \end{aligned}$$

Schließlich lassen sich die Kovarianz und Varianz mit Hilfe gewisser arithmetischer Mittel berechnen - dies wird *Verschiebungssatz* genannt:

$$\begin{aligned} \sigma(x \mid y) &= A(xy) - A(x)A(y) \\ \sigma^2(x) &= A(x^2) - A(x)^2 \end{aligned}$$

Beweis: All diese Eigenschaften lassen sich elementar nachrechnen. Zur Linearität: die erste Eigenschaft ist klar, denn $A(a) = (a + a + \cdots + a)/n = (na)/n = a$. Die beiden anderen Eigenschaften folgen genauso einfach:

$$A(ax) = \frac{1}{n}\sum_{t=1}^{n}(ax_t) = a\frac{1}{n}\sum_{t=1}^{n}x_t = aA(x)$$

$$A(x+y) = \frac{1}{n}\sum_{t=1}^{n}(x_t + y_t) = \frac{1}{n}\left(\sum_{t=1}^{n}x_t + \sum_{t=1}^{n}y_t\right) = A(x) + A(y)$$

Und mit Hilfe dieser Eigenschaften des arithmetischen Mittels lassen sich bereits die Verschiebungssätze beweisen:

$$\begin{aligned}
\sigma(x \mid y) &= A\Big((x - A(x))(y - A(y))\Big) \\
&= A\Big(xy - A(x)y - A(y)x + A(x)A(y)\Big) \\
&= A(xy) - A\Big(A(x)y\Big) - A\Big(A(y)x\Big) + A\Big(A(x)A(y)\Big) \\
&= A(xy) - A(x)A(y) - A(y)A(x) + A(x)A(y) \\
&= A(xy) - A(x)A(y) \\
\sigma^2(x) &= \sigma(x \mid x) = A(xx) - A(x)A(x) = A(x^2) - A(x)^2
\end{aligned}$$

Die Symmetrie der Kovarianz ist klar: $\sigma(x \mid y) = A(xy) - A(x)A(y) = A(yx) - A(y)A(x) = \sigma(y \mid x)$. Und mit Hilfe der Linearität und des Verschiebungssatzes folgt dann auch die Bilinearität der Kovarianz:

$$\begin{aligned}
\sigma(a \mid x) &= A(ax) - A(a)A(x) \\
&= aA(x) - aA(x) = 0 \\
\sigma(ax \mid y) &= A(axy) - A(ax)A(y) \\
&= aA(xy) - aA(x)A(y) = a\sigma(x \mid y) \\
\sigma(x + x' \mid y) &= A((x + x')y) - A(x + x')A(y) \\
&= A(xy + x'y) - (A(x) + A(x'))A(y) \\
&= A(xy) + A(x'y) - A(x)A(y) - A(x')A(y) \\
&= A(xy) - A(x)A(y) + A(x'y) - A(x')A(y) \\
&= \sigma(x \mid y) + \sigma(x' \mid y)
\end{aligned}$$

□

Zu Lagemaßen:

Bemerkung: In Kapitel 6 haben wir behauptet, dass das arithmetische Mittel und der Zentralwert durch orthogonale Projektion des $x = (x_1, x_2, \ldots, x_n)$ auf die Diagonale $\mathbb{R}1$ entstehen. Wir wollen diese Behauptung nun präzisieren und beweisen. Die Punkte auf der Diagonalen $\mathbb{R}1$ sind von der Form $a = (a, a, \ldots, a) \in \mathbb{R}^n$. Unter der Projektion versteht man, dass man dasjenige a sucht, so dass die Punkte a und x möglichst nahe beieinander liegen. In den nächsten beiden Sätzen werden wir (unter anderem) folgendes beweisen: misst man den Abstand mit $d_2(a, x)$, so kommt man beim arithmetischen Mittel $a = A(x)$ an, misst man den Abstand mit hingegen mit $d_1(a, x)$, so erreicht man den Zentralwert $a = Z(x)$.

Satz 16.5: Seien $x_1, \ldots, x_n \in \mathbb{R}$ beliebige und $\omega_1, \ldots, \omega_n \in \mathbb{R}^+$ seien positive reelle Zahlen, mit $\omega_1 + \cdots + \omega_n = 1$. Betrachten wir nun die gewichtete Abstandsfunktion

$$d \;:\; \mathbb{R} \to \mathbb{R}^+ \;:\; a \mapsto \sqrt{\sum_{i=1}^{n} \omega_i (x_i - a)^2}$$

dann hat d ein (globales) Minimum bei $a = \omega_1 x_1 + \cdots + \omega_n x_n$. Insbesondere erhalten wir das übliche arithmetische Mittel für die Gewichte $\omega_i = 1/n$.

Beweis: Da die Funktion $\mathbb{R}^+ \to \mathbb{R}^+ : y \mapsto y^2$ echt ordnungserhaltend ist, genügt es anstelle d nur $q := d^2$ zu minimieren. Wir suchen also die kritischen Punkte von $q : \mathbb{R} \to \mathbb{R}^+$ auf, d.h. wir suchen die Nullstellen der Ableitung (vergleiche z.B. [Barner, Flohr Analysis I], Kapitel 8.2)

$$\begin{aligned} 0 \;=\; q'(a) &= \sum_{i=1}^{n} \omega_i (-1) 2 (x_i - a) \\ &= (-2) \sum_{i=1}^{n} \omega_i x_i + 2a \sum_{i=1}^{n} \omega_i \;=\; 2a - 2 \sum_{i=1}^{n} \omega_i x_i \end{aligned}$$

woraus sofort $a = \omega_1 x_1 + \cdots + \omega_n x_n$ folgt. Das an dieser Stelle tatsächlich ein Minimum vorliegt folgt wie üblich, aus

$$q''(a) \;=\; (q')'(a) \;=\; 2 \;>\; 0$$

□

Satz 16.6: Es seien die Werte $x_1 \le x_2 \le \cdots \le x_n \in \mathbb{R}$ gegeben. Wir bezeichnen die Intervalle $I_0 :=]-\infty, x_1[$, bzw. für $k \in 1 \ldots (n-1)$ sei $I_k := [x_k, x_{k+1}[$ und schließlich $I_n := [x_n, \infty[$, Ferner betrachten wir die Abbildungen $D : \mathbb{R} \to \mathbb{R}^+$, $\Delta : \mathbb{R} \to \mathbb{Z}$ und $S : \mathbb{R} \to \mathbb{R}$, definiert durch

$$\begin{aligned} D(a) &:= \sum_{i=1}^{n} |x_i - a| \\ \Delta(a) &:= \#\{\, i \in 1 \ldots n \mid x_i \le a \,\} - \#\{\, i \in 1 \ldots n \mid x_i > a \,\} \end{aligned}$$

$$S(a) := \sum_{x_i > a} x_i - \sum_{x_i \leq a} x_i$$

Dann ist D stetig und positiv, S und Δ sind konstant auf den Intervallen I_0 bis I_n und D ist auf diesen Intervallen affin-linear. Ferner gilt für alle $a \in \mathbb{R}$

$$D(a) \quad = \quad S(a) + a\Delta(a)$$

Und bezeichnen wir $h := n/2$ für n gerade bzw. $h := (n+1)/2$ für n ungerade, dann hat D ein (nicht zwingend eindeutiges) absolutes Minimum in x_h.

Bemerkung: Der Beweis ist elementar, aber erstaunlich aufwendig. Die zentrale Einsicht besteht darin, dass D stückweise affin-linear ist. Ein Minimum kann also nur in einem Knick (einem der x_i) oder in einer horizontal verlaufenden Linie liegen. Mit etwas Geschick lässt sich zeigen, dass horizontal verlaufende Linien nur bei den mittleren x_i vorkommen können. Man muss also noch argumentieren, dass x_h eines der Minima ist.

Beweis:

1. Natürlich ist $a \mapsto x_i - a$ affin-linear, insbesondere stetig. Also ist auch $a \mapsto |x_i - a|$ stetig, als Verkettung stetiger Funktionen. Schließlich ist D stetig, als Summe stetiger Funktionen.

2. Die Identität $D(a) = S(a) + a\Delta(a)$ folgt durch elementare Rechung

$$\begin{aligned} D(a) &= \sum_{i=1}^{n} |x_i - a| = \sum_{x_i \leq a} (a - x_i) + \sum_{x_i > a} (x_i - a) \\ &= \sum_{x_i \leq a} a - \sum_{x_i \leq a} x_i + \sum_{x_i > a} x_i - \sum_{x_i > a} a \\ &= \sum_{x_i > a} x_i - \sum_{x_i \leq a} x_i + a \left(\sum_{x_i \leq a} 1 - \sum_{x_i > a} 1 \right) \\ &= S(a) + a\Delta(a) \end{aligned}$$

3. Ist $i \in 1 \ldots n$ und $a \in I_k$ $(k \in 1 \ldots n)$, dann ist $x_i \leq a \iff x_i < x_{k+1}$ [denn $x_i \leq a$ und $a < x_{x+1}$ implizieren $x_i < x_{k+1}$ und umgekehrt folgt aus $x_i < x_{k+1}$ aufgrund der Anordnung $i \leq k$ und damit $x_i \leq x_k \leq a$]. Für $k = 0$ ist diese Äquivalenz trivial (beide Aussagen sind falsch). Sind also a und $b \in I_k$ so folgt

$$x_i \leq a \iff x_i < x_{k+1} \iff x_i \leq b$$

Die Aussagen $x_i \leq a$ und $x_i \leq b$ sind also äquivalent und damit sind $\Delta(a) = \Delta(b)$ und $S(a) = S(b)$. Sprich Δ und S sind jeweils auf den Intervallen I_k konstant. Aufgrund von (2) ist damit auch klar, dass D auf diesen Intervallen jeweils affin-linear ist.

4. Für $a < x_1$ ist offensichtlich $\Delta(a) = -n$ und $S(a) = X := x_1 + \cdots + x_n$. Nach (2) also $D(a) = X - na$. Damit fällt D streng monoton auf I_0. Für $a > x_n$ hingegen ist $\Delta(a) = n$ und $S(a) = -X$, also $D(a) =$

$-X + na$. Sprich D ist auf I_n streng monoton steigend. Nach (1) ist D stetig, nimmt also auf dem Kompaktum $[x_1, x_n]$ ein Minimum an (siehe [Barner, Flohr, Analysis I] Kapitel 7.3). Sei nun $m \in [x_1, x_n]$ ein solches Minimum, d.h. für alle $a \in [x_1, x_n]$ gelte $D(a) \geq D(m)$. Ist nun $b > 0$, dann folgt

$$D(x_1 - b) \;=\; X - n(x_1 - b) \;=\; D(x_1) + nb \;>\; D(x_1) \;\geq\; D(m)$$

$$D(x_n + b) \;=\; -X + n(x_n + b) \;=\; D(x_n) + nb \;>\; D(x_n) \;\geq\; D(m)$$

Es ist also $D(a) \geq D(m)$ auch für a außerhalb von $[x_1, x_n]$. Sprich das Minimum in $[x_1, x_n]$ ist bereits ein absolutes Minimum von D.

5. Seien nun $1 \leq r \leq s \leq n$ dann werden wir folgende Identität benötigen

$$\begin{aligned} S(x_r) - S(x_s) = & \sum_{x_i > x_r} x_i - \sum_{x_i \leq x_r} - \sum_{x_i > x_s} x_i + \sum_{x_i \leq x_s} x_i \\ = & \left(\sum_{x_i > x_r} x_i - \sum_{x_i > x_s} x_i \right) + \left(\sum_{x_i \leq x_s} x_i - \sum_{x_i \leq x_r} \right) \\ = & \sum_{x_r < x_i \leq x_s} x_i + \sum_{x_r < x_i \leq x_s} x_i \;=\; 2 \sum_{x_r < x_i \leq x_s} x_i \end{aligned}$$

6. Sei nun $k := \#\{ x_1, \dots, x_n \}$ die Zahl der verschiedenen Werte unter den x_i und seinen $u_1 < u_2 < \dots < u_k$ eben diese Werte (d.h. für die Mengen gilt $\{ u_1, \dots, u_k \} = \{ x_1, \dots, x_n \}$). Dann setzen wir schließlich noch $N_j := \#\{ i \in 1 \dots n \mid x_i = u_j \}$. Und damit folgt dann

$$\Delta(u_j) \;=\; \big(N_1 + \dots + N_j\big) - \big(N_{j+1} + \dots + N_k\big)$$

7. Sei nun $1 \leq r < n$ mit $x_r < x_{r+1}$, dann wählen wir $j \in 1 \dots k$ so, dass $u_{j-1} = x_r$. Und damit ist auch klar, dass $u_j = x_{r+1}$ ist. Dann ist

$$\begin{aligned} D(x_r) - D(x_{r+1}) \;&=\; D(u_{j-1}) - D(u_j) \\ &=\; S(u_{j-1}) - S(u_j) + u_{j-1}\Delta(u_{j-1}) - u_j \Delta(u_j) \end{aligned}$$

Und aus (5) erhalten wir offenbar aber $S(u_{j-1}) - S(u_j) = 2u_j N_j$, also

$$\begin{aligned} & D(x_r) - D(x_{r+1}) \\ = \;& 2u_j N_j + u_{j-1}\Delta(u_{j-1}) - u_j \Delta(u_j) \\ = \;& 2u_j N_j + u_{j-1} \left(\sum_{i=1}^{j-1} N_i - \sum_{i=j}^{k} N_i \right) \\ & - u_j \left(\sum_{i=1}^{j} N_i - \sum_{i=j+1}^{k} N_i \right) \\ = \;& 2u_j N_j + u_{j-1} \sum_{i=1}^{j-1} N_i - u_{j-1} N_j - u_{j-1} \sum_{i=j+1}^{k} N_i \end{aligned}$$

$$
\begin{aligned}
& - u_j \sum_{i=1}^{j-1} N_i - u_j N_j + u_j \sum_{i=j+1}^{k} N_i \\
= \; & (u_{j-1} - u_j) \left(\sum_{i=1}^{j-1} N_i - \sum_{i=j+1}^{k} N_i + N_j \right) \\
= \; & (u_{j-1} - u_j) \left(\sum_{i=1}^{j-1} N_i - \sum_{i=j}^{k} N_i \right) \\
= \; & (u_{j-1} - u_j) \Delta(u_{j-1}) \; = \; (x_r - x_{r+1}) \Delta(x_r)
\end{aligned}
$$

8. Nach (4) besitzt D ein absolutes Munimum $m \in [x_1, x_n]$. Angenommen es ist $m \notin \{ x_1, \dots, x_n \}$. Wählen wir $h \in 1 \dots n$ maximal, mit $x_h \leq m$ so ist $m \in]x_h, x_{h+1}[\; \subseteq \; I_h$. Nach (3) ist D in einer Umgebung von m affin-linear und damit differenzierbar. Also ist

$$
\begin{aligned}
0 \; & = \; D'(m) \; = \; \Delta(m) \; = \; \Delta(x_h) \\
& = \; \# \{ i \in 1 \dots n \mid x_i \leq x_h \} - \# \{ i \in 1 \dots n \mid x_i > x_h \} \\
& = \; h - (n - h) \; = \; 2h - n
\end{aligned}
$$

Mithin ist n gerade und $h = n/2$. Ferner gilt wegen $\Delta(x_h) = 0 = \Delta(m)$ nach (3) auch $D(x_h) = S(x_h) = S(x_m) = D(m)$. Sprich x_h ist ebenfalls ein absolutes Minimum von D.

9. Nach (8) genügt es also noch den Fall $m \in \{ x_1, \dots, x_n \}$ zu betrachten. Wir bezeichnen u_j und N_j wie zuvor in (7) und wählen $j \in 1 \dots k$ so, dass $u_j = m$. Da $D(m)$ minimal ist, folgt $D(u_j) \leq D(u_{j+1})$ und damit (unter Verwendung von (7)) $0 \geq D(u_j) - D(u_{j+1}) = (u_j - u_{j+1}) \Delta(u_j)$. Und wegen $u_j - u_{j+1} < 0$ haben wir also $\Delta(u_j) \geq 0$, das bedeutet

$$
\sum_{i=1}^{j} N_i \; \geq \; \sum_{i=j+1}^{k} N_i
$$

Genauso ist $D(u_{j-1}) \geq D(u_j)$ und damit $0 \leq D(u_{j-1}) - D(u_j) = (u_{j-1} - u_j) \Delta(u_{j-1})$ also $\Delta(u_{j-1}) \leq 0$ was bedeutet, dass

$$
\sum_{i=1}^{j-1} N_i \; \leq \; \sum_{i=j}^{k} N_i
$$

Setzen wir $A := N_1 + \dots + N_{j-1}$ und $B := N_{j+1} + \dots + N_k$ so haben wir also die Gleichungen erhalten $A + N_j \geq B$ und $A \leq N_j + B$, wobei

$$
\underbrace{x_1 \leq \dots \leq x_A}_{A} < \underbrace{x_{A+1} = \dots = x_{A+N_j}}_{N_j} < \underbrace{x_{A+N_j+1} \leq \dots \leq x_n}_{B}
$$

Nun setzen wir $h := n/2$ für n gerade und $h := (n+1)/2$ für n ungerade. Dann erhalten wir $2A = A + A \leq A + N_j + B = n$ und damit $A \leq n/2 \leq h$. Andererseits ist $2B = B + B \leq A + N_j + B = n$ und damit auch $B \leq n/2$. Daraus folgt $A + N_j = n - B \geq n - n/2 = n/2$. Da $A + N_j$ aber eine natürliche Zahl (und h die kleinste natürliche Zahl über $n/2$) ist folgt daraus auch $A + N_j \geq h$. Insgesamt haben wir

$$A \;\leq\; h \;\leq\; A + N_j$$

1. Fall: ist $A < h$, dann ist $h \in (A+1) \dots (A + N_j)$ und damit $x_h = u_j = m$. Sprich x_h ist ein absolutes Minimum von D. 2. Fall: ist $A = h$ dann haben wir $n/2 \leq h = A \leq n/2$ und damit $A = n/2$. Damit folgt auch $N_j + B = n - A = n/2 = A$. Nun bedeutet $A = N_j + B$ aber wiederum $\Delta(u_{j-1}) = 0$ und damit $D(u_{j-1}) - D(u_j) = (u_{j-1} - u_j)\Delta(u_{j-1}) = 0$. Wegen $u_{j-1} = x_A = x_h$ und $m = u_j$ folgt daraus also $D(x_h) = D(u_{j-1}) = D(u_j) = D(m)$. Also ist x_h auch in diesem Fall ein absolutes Minimum von D.

□

Zum Gini-Koeffizienten:

Satz 16.7: Seien die Werte $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \in \mathbb{R}$ gegeben und bezeichne $X := x_1 + x_2 + \cdots + x_n$ deren Summe. Dann gilt für beliebiges $k \in 1 \dots n$ stets die Abschätzung

$$\frac{x_1 + \cdots + x_k}{X} \;\leq\; \frac{k}{n}$$

Beweis: Um dies zu sehen definieren wir die Summen $a := x_1 + \cdots + x_k$ und $b := x_{k+1} + \cdots + x_n$. Dann ist also $X = a + b$ und da für $i \leq k$ stets $x_i \leq x_k$ gilt, haben wir auch $a \leq k x_k$. Und entsprechend wegen $x_j \geq x_{k+1}$ für $j \geq k+1$ ist auch $b \geq (n-k) x_{k+1}$. Damit haben wir dann $a \leq k x_k \leq k x_{k+1} \leq \frac{k}{n-k} b$. Und daraus erhalten wir unmittelbar eine weitere Abschätzung

$$\left(1 - \frac{k}{n}\right) a \;=\; \frac{n-k}{n} a \;\leq\; \frac{n-k}{n} \cdot \frac{k}{n-k} b \;\leq\; \frac{k}{n} b$$

Also $a - (k/n)a \leq (k/n)b$ und damit $a \leq (k/n)(a+b) = (k/n)X$. Durch Division mit X erhalten wir also $(x_1 + \cdots + x_k)/X = a/X \leq k/n$.

□

Bemerkung 16.8: Wie immer ist $x = (x_1, x_2, \dots, x_n)$ und wir bezeichnen die *verschiedenen* Zahlen unter den x_t mit r_i und die *absolute* Häufigkeit von r_i mit $n_i := \#\{\, t \in 1 \dots n \mid x_t = r_i \,\}$. Die *relative* Häufigkeit von r_i ist wieder $h_i := n_i/n$. Dann haben wir auch die *absolute Merkmalssumme* $X_i := n_i r_i$ von r_i eingeführt, ebenso die *totale Merkmalssumme*

$$X \;=\; \sum_{i=1}^{p} X_i \;=\; \sum_{i=i}^{p} n_i r_i \;=\; \sum_{t=1}^{n} x_t \;=\; nA(x)$$

für die Konzentrationsanalyse sind auch die *relativen Merkmalssummen* $\ell_i := X_i/X = (n_r r_i)/X = (h_i r_i)/A(x)$ von r_i interessant. Und wenn man diese aufsummiert, erhält man die *kumulierte, relative Merkmalssumme*

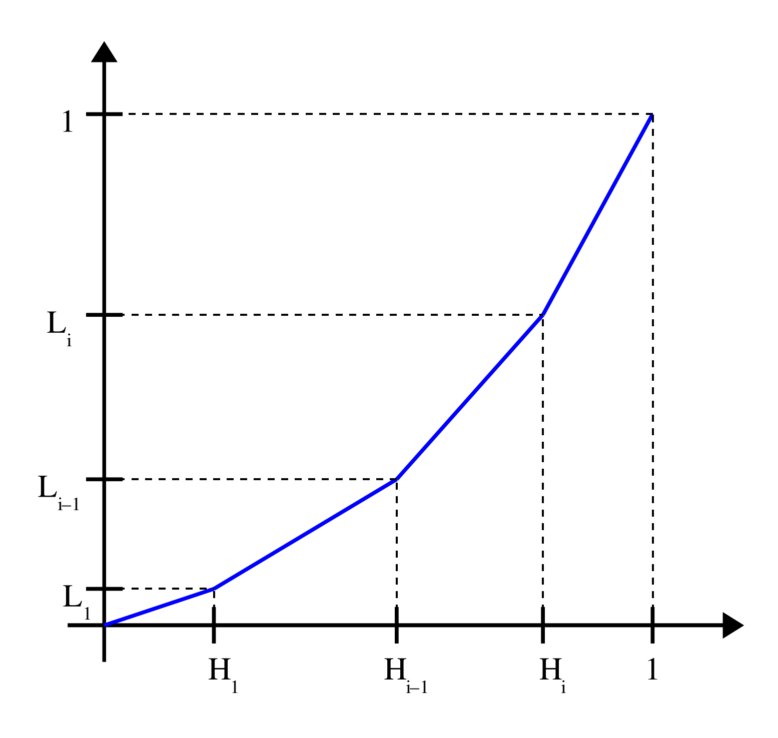
$$L_k \;:=\; \sum_{i=1}^{k} \ell_i \;=\; L_{k-1} + \ell_k$$

Nach dem obigen Satz gilt insbesondere $L_k \leq N_k/n = H_k$, d.h. die Lorenz-Kurve $L : [0,1] \to [0,1]$ hängt tatsächlich unter der Diagonalen durch. Wir werden im folgenden zwei Formeln zur Berechnung des Gini-Koeffizienten geben und beweisen.

Satz 16.9: Mit den Bezeichnungen aus der Bemerkung oben, lässt sich der Gini-Koeffizient von $x = (x_1, x_2, \dots, x_n)$ mit Hilfe der kumulierten, relativen Merkmalssummen berechnen:

$$R \;=\; 1 - \sum_{i=1}^{p} h_i (L_{i-1} + L_i)$$

Beweis:



Wir werten zunächst einmal die Fläche unter der Lorenz-Kurve aus. Dazu zerlegen wir diese Fläche in Trapeze zwischen $t = H_{i-1}$ und $t = H_i$. Diese haben also die Breite $Hi - H_{i-1} = h_i$ und die mittlere Höhe $(L_{i-1} + L_i)/2$ (siehe Abbildung). Insgesamt also:

$$A \;:=\; \int_0^1 L(t)dt \;=\; \sum_{i=1}^{p} \text{Trapezfläche}_i \;=\; \sum_{i=1}^{p} h_i \frac{L_{i-1} + L_i}{2}$$

Der Gini-Koeffizient ist nach Definition das 2fache der Fläche zwischen der Lorenz-Kurve A und der Diagonalen 1/2. Wir erhalten also die Behauptung

$$R \;=\; 2\left(\frac{1}{2} - A\right) \;=\; 1 - 2\sum_{i=1}^{p} h_i \frac{L_{i-1} + L_i}{2} \;=\; 1 - \sum_{i=1}^{p} h_i L_{i-1} + L_i$$

□

Satz 16.10: Seien die Werte $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \in \mathbb{R}$ gegeben und bezeichne $X := x_1 + x_2 + \cdots + x_n$ deren Summe. Dann ist der Gini-Koeffizient dieser Werte gegeben, durch

$$R \;=\; \frac{2}{nX}\left(\sum_{i=1}^{n} i x_i\right) - \frac{n+1}{n}$$

Beweis: Wir bezeichnen $u_k := k/n$ und $v_k := (x_1 + \cdots + x_k)/X$. Und aus formalen Gründen setzen wir auch $u_0 := 0$ und $v_0 := 0$. Ferner sei $L : [0,1] \to [0,1]$ die Lorenz-Kurve der Werte x_i. Nach Definition der Lorenz Kurve ist $L(t)$ für $t \in [u_{k-1}, u_k]$ gerade gegeben, durch

$$L(t) \;=\; \frac{t - u_{k-1}}{u_k - u_{k-1}}(v_k - v_{k-1}) + v_{k-1} \;=\; (nt - k)\frac{x_k}{X} + v_k$$

Die Fläche unter der Lorenz-Kurve lässt sich also Aufspalten in die Trapeze auf den Intervallen $[u_{k-1}, u_k]$ berechnen. Deren Breite beträgt $u_k - u_{k-1} = 1/n$ und deren mittlere Höhe ist $(v_{k-1} + v_k)/2$. Die Fläche wird also zu

$$\begin{aligned}
A \;&=\; \int_0^1 L(t)dt \;=\; \sum_{k=1}^{n} \int_{u_{k-1}}^{u_k} L(t)dt \\
&=\; \sum_{k=1}^{n} \left(\frac{v_{k-1} + v_k}{2}\right)(u_k - u_{k-1}) \\
&=\; \sum_{k=1}^{n} \frac{2x_1 + \cdots + 2x_{k-1} + x_k}{2X} \cdot \frac{1}{n} \\
&=\; \frac{1}{nX}\sum_{k=1}^{n}\left(\sum_{i=1}^{k-1} x_i + \frac{x_k}{2}\right)
\end{aligned}$$

Der Gini-Koeffizient ist definiert, als 2mal die Fläche zwischen der Diagonalen und der Lorenz-Kurve. Da die Lorenz-Kurve (wie gesehen) stets unter der Diagonalen liegt, lässt er sich also wie folgt berechnen

$$
\begin{aligned}
R &= 2\left(\frac{1}{2} - A\right) = 1 - 2A = 1 - \frac{1}{nX}\sum_{k=1}^{n}\left(2\sum_{i=1}^{k-1} x_i + x_k\right) \\
&= 1 - \frac{1}{nX}\left[x_1 + (2x_1 + x_2) + \cdots + (2x_1 + \cdots + 2x_{n-1} + x_n\right] \\
&= 1 - \frac{1}{nX}\sum_{i=1}\left[2(n-i)+1\right]x_i = 1 - \frac{1}{nX}\sum_{i=1}\left[(2n+1) - 2i\right]x_i \\
&= 1 - \frac{1}{nX}\left((2n+1)X + 2\sum_{i=1}^{n} ix_i\right) = 1 - \frac{2n+1}{n} + \frac{2}{nX}\sum_{i=1}^{n} ix_i \\
&= -\frac{n+1}{n} + \frac{2}{nX}\sum_{i=1}^{n} ix_i
\end{aligned}
$$

□

Zur Regression:

In diesem Abschnitt untersuchen wir den Zusammenhang zwischen den Größen x und $y \in \mathbb{R}^n$. Zu jedem $t \in 1 \ldots n$ haben wir also einen Datenpunkt $(x_t \mid y_t)$ gegeben. Im einfachsten Fall folgen die Punkte einem linearen Zusammenhang $y = ax + b$. Es geht also darum die Parameter a und b so anzupassen, dass sich die Gerade $ax + b$ möglichst dicht an die Datenpunkte anschmiegt. Dieses Problem wird im nächsten Satz präzisiert und gelöst werden. Danach werden wir das allgemeinere Problem eines polynomialen Zusammenhangs $y = a_m x^m + \cdots + a_1 x + a_0$ lösen. Um dies zu bewältigen ist aber ein wenig mehr, als Schulmathematik, notwendig.

Satz 16.11: Seien $x = (x_1, x_2, , x_n)$ und $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$. Dann suchen wir die *Regressionsgerade*, $g(x) = ax + b$, die den folgenden Abstand minimiert:

$$
d(a,b) = \sqrt{\sum_{t=1}^{n}\left(ax_t + b - y_t\right)^2}
$$

Enthält x mindestens 2 verschiedene Werte, so lässt sich dieses Problem eindeutig lösen, durch

$$
\begin{aligned}
a &= \frac{\sigma(x \mid y)}{\sigma(x)^2} \\
b &= A(y) - aA(x)
\end{aligned}
$$

Beweis: Da die Funktion $\mathbb{R}^+ \to \mathbb{R}^+ : x \mapsto x^2$ echt ordnungserhaltend ist, genügt es die Funktion $q(a,b) := d(a,b)^2$ zu minimieren. Ferner bezeichne $X := \sum_t x_t = nA(x)$ und $Y := \sum_t y_t = nA(y)$ die Summe der x_t bzw. y_t. Dann rechnet man nach, dass

$$\begin{aligned} q(a,b) &= \sum_{t=1}^{n} \big(ax_t + b - y_t\big)^2 \\ &= \sum_{t=1}^{n} \big(a^2x_t^2 + b^2 - y_t^2\big)^2 - 2\sum_{t=1}^{n} \big(abx_t - ax_ty_t - by_t\big)^2 \\ &= \langle x \mid x\rangle a^2 + nb^2 + \langle y \mid y\rangle + 2Xab - 2\langle x \mid y\rangle a - 2Yb \end{aligned}$$

Damit $q(a,b)$ minimal wird, müssen die Ableitungen nach a *und* nach b verschwinden. Die Ableitung nach b liefert $0 = \partial_b q(a,b) = 2nb + 2Xa - 2Y$, also $Xa + nb = Y$. Dividieren wir diese Gleichung noch durch n erhalten wir

$$\partial_b q(a,b) = 0 \quad \Longrightarrow \quad A(x)a + b = A(y)$$

Die Ableitung nach a liefert $0 = \partial_a q(a,b) = 2\langle x \mid x\rangle a + 2Xb - 2\langle x \mid y\rangle$, also $\langle x \mid x\rangle a + Xb = \langle x \mid y\rangle$. Dividieren wir auch diese Gleichung durch n erhalten wir

$$\partial_a q(a,b) = 0 \quad \Longrightarrow \quad A(x^2)a + A(x)b = A(xy)$$

Wir haben bereits (aus der ersten Gleichung) $b = A(y) - A(x)a$. Setzen wir dies in die zweite Gleichnung ein, so erhalten wir

$$\begin{aligned} A(x^2)a + A(x)\big(A(y) - A(x)a\big) &= A(xy) \\ \big(A(x^2) - A(x)^2\big)a &= A(xy) - A(x)A(y) \end{aligned}$$

Setzen wir die Verschiebungssätze ein so haben wir also $\sigma(x)^2 a = \sigma(x \mid y)$ erhalten. Nach Voraussetzung sind die x_t nicht konstant, also $\sigma(x) \neq 0$. Durch Division durch $\sigma(x)^2$ erhalten wir damit die Behauptung

$$a = \frac{\sigma(x \mid y)}{\sigma(x)^2}$$

Dass die Funktion $q(a,b)$ tatsächlich ein Minimum an dieser Stelle (a,b) hat, erkennt man an der Determinante der Hesse-Matrix (siehe [Friedberg, Insel, Spence, Linear Algebra] Theorem 6.31)

$$\partial_a\partial_a q(a,b) \cdot \partial_b\partial_b q(a,b) - \big(\partial_a\partial_b q(a,b)\big)^2 = \ldots = 4n^2\sigma(x)^2 > 0$$

□

Bemerkung 16.12: Das Problem der linearen Regression wurde also gelöst, indem man den (mit d_2 gemessenen) Abstand zwischen den Punkten $(g(x_1), g(x_2), \ldots, g(x_n))$ und $y \in \mathbb{R}^n$ minimiert hat. Dies führte auf die *Gauss'schen Normalengleichungen*

$$\begin{aligned} A(x)a + b &= A(y) \\ A(x^2)a + A(x)b &= A(xy) \end{aligned}$$

Die erste dieser beiden Gleichungen enthält, dass die *Regressionsgerade* $g(x) = ax + b$ durch den Punkt $(A(x) \mid A(y)) \in \mathbb{R}^2$ der arithmetischen Mittel läuft

$$A(y) \ = \ A(x)a + b \ = \ g(A(x))$$

Neben dieses geometrische Bild gesellt sich aber auch ein statistisches Bild: bezeichnen wir die Abweichung bei der linearen Approximation mit

$$u_t \ := \ y_t - g(x_t)$$

Dann stellt man zunächst fest, dass die mittlere Abweichung $A(u)$ bei der linearen Regression verschwindet, denn man rechnet nach, dass

$$A(u) \ = \ A(y - (ax + b)) \ = \ A(y) - \big(aA(x) + b\big) \ = \ 0$$

Damit ergibt sich dann, dass - indem wir $d(a, b)$ minimiert haben - gleich auch die Varianz $\sigma(u)^2$ minimiert wurde, denn man sieht

$$\sigma(u)^2 \ = \ A(u^2) - A(u)^2 \ = \ A(u^2) \ = \ \frac{1}{n} d(a, b)^2$$

Satz 16.13: Seien wieder $x, y \in \mathbb{R}^n$ und bezeichne $y = ax+b$ die zugehörige Regressionsgerade. Sei weiterhin $u_t = y_t - (ax_t + b)$ die Abweichung von y_t von der Geraden. Dann besteht die folgende *Streuungszerlegung*

$$\sigma(y)^2 \ = \ \sigma(ax + b)^2 + \sigma(u)^2$$

Und für den *Korrelationskoeffizienten* r gilt schließlich noch die Beziehung

$$r \ := \ \frac{\sigma(x \mid y)}{\sigma(x)\sigma(y)} \quad \Longrightarrow \quad r^2 \ = \ \frac{\sigma(ax + b)^2}{\sigma(y)^2}$$

Bemerkung 16.14: Der obige Satz sagt anschaulich, dass sich die Streuung von y aus zwei Teilen zusammen setzt: (1) durch den Zusammenhang $y \approx ax + b$ wird die Streuung der x zu einer Streuung der y übertragen. Dies ist also der *erklärte Teil* der Streuung. (2) die *Reststreuung* die von der Abweichung herrührt und nicht erklärt wird. Der Korrelationskoeffizient r misst also den erklärten Anteil an der Streuung der y und ist somit und ist somit ein Maß für die Glaubwürdigkeit des Zusammenhangs.

Beweis:

1. Der Übersichtlichkeit halber zerlegen wir den Beweis in mehrere Teilschritte. Als ersten Schritt beweisen wir eine Aussage über $ax+b \in \mathbb{R}^n$

$$A\big((ax+b)^2\big) \;=\; A\big((ax+b)y\big)$$

Denn $A((ax+b)^2) = A(a^2x^2 + 2abx + b^2) = a^2A(x^2) + 2abA(x) + b^2$. Diesen Ausdruck spalten wir geschickt auf und wenden die Gauss'schen Normalengleichungen (16.12) an

$$\begin{aligned} A\big((ax+b)^2\big) &= a\big(aA(x^2) + bA(x)\big) + b\big(aA(x) + b\big) \\ &= aA(xy) + bA(y) \\ &= A\big((ax+y)y\big) \end{aligned}$$

2. Mit Hilfe von (1) und der Tatsache dass $A(u) = 0$ ist, können wir dann auch die Varianz der Abweichungen $u \in \mathbb{R}^n$ mit dem Verschiebungssatz (16.4) umrechnen, zu

$$\begin{aligned} \sigma(u)^2 &= A\big((y - (ax+b))^2\big) \\ &= A\big(y^2 - 2(ax+b)y + (ax+b)^2\big) \\ &= A(y^2) - 2A((ax+b)y) + A((ax+b)^2) \\ &= A(y^2) - 2A((ax+b)^2) + A((ax+b)^2) \\ &= A(y^2) - A((ax+b)^2) \end{aligned}$$

3. Nachdem wir in (2) die Varianz von u berechnet haben, können wir mit Hilfe von $A(ax+b) = aA(x) + b = A(y)$ die Streuungszerlegung recht leicht nachrechnen:

$$\begin{aligned} \sigma(y)^2 - \sigma(u)^2 &= A(y^2) - A(y)^2 - A(y^2) + A((ax+b))^2 \\ &= A((ax+b))^2 - A(y)^2 \\ &= A((ax+b))^2 - A(ax+b)^2 \\ &= \sigma(ax+b)^2 \end{aligned}$$

4. Als nächstes benötigen wir wieder eine Hilfsaussage für die Kovarianz zwischen x und y, dieses Mal

$$\sigma(x \mid y) \;=\; \sigma(x \mid ax+y)$$

Denn setzen wir die Gauss'schen Normalengleichungen auf den Verschiebungssatz der Kovarianz an, so finden wir

$$\begin{aligned} \sigma(x \mid y) &= A(xy) - A(x)A(y) \\ &= \big(aA(x^2) + bA(x)\big) - A(x)A\big(ax+b\big) \\ &= A\big(x(ax+b)\big) - A(x)A\big(ax+b\big) \\ &= \sigma(x \mid ax+b) \end{aligned}$$

5. Und mit Hilfe von (4) und der Streuungszerlegung können wir schließlich auch die Identität für den Korrelationskoeffizienten nachrechen:

$$\begin{aligned}
\sigma(y)^2 r^2 &= \sigma(y)^2 \frac{\sigma(x \mid y)^2}{\sigma(x)^2 \sigma(y)^2} \\
&= \frac{\sigma(x \mid y)}{\sigma(x)^2} \sigma(x \mid y) \\
&= a\sigma(x \mid y) \\
&= a\sigma(x \mid ax + b) \\
&= \sigma(ax + b \mid ax + b) - \sigma(b \mid ax + b) \\
&= \sigma(ax + b \mid ax + b) \; = \; \sigma(ax + b)^2
\end{aligned}$$

Die letzte Gleichung besteht, da durch die Konstanz von $b = (b, b, \dots, b) \in \mathbb{R}^n$ die Kovarianz $\sigma(b \mid ax + b) = 0$ verschwindet (siehe (16.4)).

□

Satz 16.15: Seien Ω eine beliebige Menge, $1 \leq m$, $n \in \mathbb{N}$ und weiterhin $x = (x_1, x_2, \dots, x_n) \in \Omega^n$ und $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ beliebig. Zu jedem $i \in 1 \dots m$ sei ferner je eine Funktion $f_i : \Omega \to \mathbb{R}$ gegeben. Damit definieren wir dann die $(n \times m)$-Matrix

$$A \; := \; \begin{pmatrix} f_1(x_1) & \cdots & f_m(x_1) \\ \vdots & & \vdots \\ f_1(x_n) & \cdots & f_m(x_n) \end{pmatrix}$$

Mit A^* bezeichnen die transponierte Matrix von A. Sei nun weiterhin das Tupel $a = (a_1, a_2, \dots, a_m) \in \mathbb{R}^m$ vorgelegt, dann definieren wir die Funktion f_a und führen damit auch die Abstandsfunktion d ein

$$f_a \; : \; \Omega \to \mathbb{R} \; : \; x \mapsto \sum_{k=1}^{m} a_k f_k(x)$$

$$d \; : \; \mathbb{R}^m \to \mathbb{R} \; : \; a \mapsto \sqrt{\sum_{i=1}^{n} (f_a(x_i) - y_i)^2}$$

Besitzt d ein Minimum in a, so sind notwendiger Weise auch die *Gauss'schen Normalengleichungen* erfüllt, die da lauten

$$A^* A\, a \; = \; A\, y$$

Beweis: Da die Funktion $\mathbb{R}^+ \to \mathbb{R}^+ : x \mapsto x^2$ echt ordnungserhaltend ist, genügt es anstelle d nur $q := d^2$ zu minimieren. Notwendig für ein Minimum von q in a ist, dass a ein kritischer Punkt von q ist (vergleiche [Barner, Flohr, Analysis II], Kapitel 14.4). Das heißt für alle $j \in 1 \dots m$ gilt

$$\begin{aligned} 0 \;=\; \partial_j q(a) &= \sum_{i=1}^{n} \partial_j \left(\sum_{k=1}^{m} a_k f_k(x_i) - y_i \right)^2 \\ &= \sum_{i=1}^{n} 2 f_j(x_i) \left(\sum_{k=1}^{m} a_k f_k(x_i) - y_i \right) \\ &= 2 \sum_{k=1}^{n} a_k \sum_{i=1}^{n} f_j(x_i) f_k(x_i) - 2 \sum_{i=1}^{n} f_j(x_i) y_i \end{aligned}$$

Nach Definition ist $a_{i,k} = f_k(x_i)$, also $a^*_{j,i} = f_j(x_i)$. Mit etwas Konzentration sieht sieht man damit, dass der r-s-te Koeffizient der $(m \times m)$-Matrix A^*A gegeben ist, durch

$$[A^*A]_{j,k} \;=\; \sum_{i=1}^{n} f_j(x_i) f_k(x_i)$$

Wir können dies in die obige Gleichung einsetzen und erhalten damit nun

$$[A\,y]_j \;=\; \sum_{i=1}^{n} f_j(x_i) y_i \;=\; \sum_{k=1}^{n} a_k \, [A^*A]_{j,k} \;=\; [A^*A\,a]_j$$

Da dies aber für alle $j \in 1 \dots m$ galt, stimmt die Gleichung $A^*A\,a = A\,y$ also auch im Sinne von Tupeln. Dies war aber gerade die Behauptung.

□

Satz 16.16: Seien nun die Potenzfunktionen $f_i : \mathbb{R} \to \mathbb{R} : x \mapsto x^i$ gegeben (wobei $i \in 0 \dots m$). Ist $a = (a_0, a_1, \dots, a_m) \in \mathbb{R}^{m+1}$ dann ist f_a also gerade das Polynom $f_a(x) = a_m x^m + \dots + a_1 x + a_0 \in \mathbb{R}[x]$. Sind nun noch die $x_i \in \mathbb{R}$ (wobei $i \in 1 \dots n$) paarweise verschieden und ist $m < n$, dann ist die $((m+1) \times (m+1))$-Matrix A^*A stets invertierbar und damit sind die Gauss'schen Normalengleichungen stets lösbar. Es ist also

$$a \;=\; (A^*A)^{-1} A\,y$$

Beweis: In einem ersten Schritt zeigen wir, dass A injektiv ist: nehmen wir also an, a läge im Kern von A, das hieße

$$0 \;=\; A\,a \;=\; \begin{pmatrix} a_0 + a_1 x_1 + \dots + a_m x_1^m \\ \vdots \\ a_0 + a_1 x_n + \dots + a_m x_n^m \end{pmatrix} \;=\; \begin{pmatrix} f_a(x_1) \\ \vdots \\ f_a(x_n) \end{pmatrix}$$

Das Polynom f_a hat also die Nullstellen x_1 bis x_n. Nach Annahme sind diese verschieden, so dass f_a mindestens n Nullstellen hat. Andererseits hat f_a höchstens den Grad m. Insgesamt erhalten wir also

$$\deg(f_a) \;\le\; m \;<\; n \;\le\; \#\{\, x \in \mathbb{R} \mid f_a(x) = 0 \,\}$$

Damit hat f_a mehr Nullstellen als sein Grad erlaubt und es folgt $f_a = 0$ (siehe [Adkins, Weintraub, Algebra] Corollary (2.4.7)). Das heißt aber nichts anderes, als $a = 0 \in \mathbb{R}^{m+1}$ und damit ist A injektiv. Im nächsten Schritt zeigen wir, dass A^*A positiv definit ist: zunächst ist A^*A symmetrisch (klar) und positiv, wegen

$$\langle A^*A\,a \mid a\rangle \;=\; \langle A\,a \mid A\,a\rangle \;= \|A\,a\|^2 \geq\; 0$$

Angenommen $\langle A^*A\,a \mid a\rangle = 0$, dann wäre also $\| A\,a\| = 0$ und damit selbst $A\,a = 0$. Wie gesehen folgt daraus aber wiederum $a = 0$ und damit ist A^*A sogar definit. Insbesondere ist

$$\det(A^*A) \;>\; 0$$

[klar: A^*A ist symmetrisch, also diagonalisierbar (vergleiche [Friedberg, Insel, Spence, Linear Algebra], Corollary to Theorem 6.29). Die Determinante ist also das Produkt der Eigenwerte. Und wegen $\langle A^*A\,a \mid a\rangle > 0$ sind alle Eigenwerte echt positiv]. Mithin ist A^*A invertierbar.

□

Bemerkung 16.17: Der vorangegangene Satz erlaubt also nicht nur lineare Regressionen auszuführen, sondern Regressionen mit Polynomen beliebigen Grades. Sind also die (paarweise verschiedenen) Stützstellen $x_1, \dots, x_n \in \mathbb{R}$ und zugehörigen Höhen $y_1, \dots, y_n \in \mathbb{R}$ gegeben, so gibt es ein eindeutig bestimmtes Polynom $f(x) = a_m x^m + \dots + a_1 x + a_0 \in \mathbb{R}[x]$ (wobei $m < n$), das den Abstand

$$d(a) \;=\; \sqrt{\sum_{i=1}^{n} (f(x_i) - y_i)^2}$$

minimiert. Und die Koeffizienten dieses Polynoms lassen sich berechnen, indem man die Gauss'schen Normalengleichungen löst, d.h. man berechnet

$$a \;=\; (A^*A)^{-1}\,A\,y$$

wobei A die folgende Matrix bezeichnet (man bemerke, dass im Fall $m = n-1$ hier gerade die Vandermond'sche Matrix zu stehen kommt)

$$A \;:=\; \begin{pmatrix} 1 & x_1 & \cdots & x_1^m \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^m \end{pmatrix}$$

Zur Zeitreihenanalyse:

Wir betrachten k Saisons $i = (1, 2, \dots, k)$. Diese Folge von Saisons wird m-mal hintereinander durchlaufen. Insgesamt betrachten wir also $n = km$ Zeitpunkte $t = (1, 2, \dots, n)$, wobei

Zeitpunkt	1	2	$\cdots$	k	$1+k$	$\cdots$	$k+k$	$1+2k$	$\cdots$	n
Saison	1	2	$\cdots$	k	1	$\cdots$	k	1	$\cdots$	k

Wir können die Zeitpunkte also durchzählen, mit $t = i + jk$ wobei $i \in 1 \dots k$ und $j \in 0 \dots (m-1)$ läuft. Dabei gibt i die Saison und j die Zahl der vergangenen Zyklen an. Zu jedem Zeitpunkt t ist nun ein Wert y_t gegeben, $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$. Dann suchen wir eine möglichst gute Näherung

$$y_{i+jk} \approx a(i+jk) + b + s_i$$

wobei a, $b \in \mathbb{R}$ und $s = (s_1, s_2, \dots, s_k) \in \mathbb{R}^k$. Doch bevor wir das Problem der besten Näherung in Satz (16.19) lösen, geben wir zunächst eine vereinfachte, etwas schlechtere Approximation an:

Satz 16.18: Sei $t = (1, 2, \dots, n)$ und $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ und bezeichne $y = at + b$ die zugehörige Regressionsgerade. Dann gilt

$$\begin{aligned} A(t) &= \frac{n+1}{2} \\ \sigma(t)^2 &= \frac{n^2-1}{12} \\ a &= \frac{12}{n^2-1} A(ty) - \frac{6}{n-1} A(y) \\ b &= A(y) - \frac{n+1}{2} a \\ \sigma(at+b)^2 &= \frac{n^2-1}{12} a^2 \end{aligned}$$

Sei nun $n = km$ und bezeichne s_i die durchschnittliche Abweichung von y_t von der Approximation $at + b$ in der Saison $i \in 1 \dots k$, d.h.

$$s_i := \frac{1}{m} \sum_{j=0}^{m-1} \Big(y_{i+jk} - \big(a(i+jk) + b \big) \Big)$$

Dann lässt sich der saisonale Einfluss s_i auch einfacher berechnen, vermöge

$$s_i = \frac{1}{m} \sum_{j=0}^{m-1} y_{i+jk} - \left(ai + b + \frac{n-k}{2} a \right)$$

Beweis: Das aritmetische Mittel folgt sofort aus den Summenformeln (16.1): $A(t) = (1/n)(n(n+1)/2) = (n+1)/2$. Und mit den Summenformeln finden wir dann weiter

$$\begin{aligned} \sigma(t)^2 &= A(t^2) - A(t)^2 = \frac{1}{n} \cdot \frac{(2n+1)(n+1)n}{6} - \left(\frac{n+1}{2} \right)^2 \\ &= (n+1) \left(\frac{2n+1}{6} - \frac{n+1}{4} \right) = (n+1) \frac{n-1}{12} = \frac{n^2-1}{12} \end{aligned}$$

Die Koeffizienten der Regressionsgeraden können wir nach (16.11) und dem Verschiebungssatz (16.4) damit berechnen, zu

$$\begin{aligned} a &= \frac{\sigma(t \mid y)}{\sigma(t)^2} = \frac{12}{n^2-1}\Big(A(ty) - A(t)A(y)\Big) \\ &= \frac{12}{n^2-1}\left(A(ty) - \frac{n+1}{2}A(y)\right) = \frac{12}{n^2-1}A(ty) - \frac{6}{n-1}A(y) \\ b &= A(y) - A(t)a = A(y) - \frac{n+1}{2}a \end{aligned}$$

Der interessanteste Teil ist die Berechnung der Varianz der Regressionsgeraden $at+b$ mit Hilfe der Linearität des arithmetischen Mittels (16.4)

$$\begin{aligned} \sigma(at+b)^2 &= A\big((at+b)^2\big) - A(at+b)^2 \\ &= A\big(a^2t^2 + 2abt + b^2\big) - \big(aA(t) + b\big)^2 \\ &= a^2A(t^2) + 2abA(t) + b^2 - a^2A(t)^2 - 2abA(t) - b^2 \\ &= a^2\big(A(t^2) - A(t^2)\big) = a^2\sigma(t)^2 = \frac{n^2-1}{12}a^2 \end{aligned}$$

Es bleibt nur noch die Formel für die saisonalen Einflüsse nachzurechnen. Dazu führen wir Y_i als Abkürzung für die Summe aller y_t, die zur Saison i gehören ein, d.h. $Y_i := \sum_j y_{i+jk}$. Damit gilt dann

$$\begin{aligned} s_i &= \frac{1}{m}\sum_{j=0}^{m-1}\Big(y_{i+jk} - \big(a(i+jk)+b\big)\Big) = \frac{1}{m}Y_i - \frac{1}{m}\sum_{j=0}^{m-1}\big(a(i+jk)+b\big) \\ &= \frac{1}{m}Y_i - ai - a\frac{k}{m}\sum_{j=0}^{m-1} j - b = \frac{1}{m}Y_i - (ai+b) - a\frac{k}{m}\frac{(m-1)m}{2} \\ &= \frac{1}{m}Y_i - (ai+b) - a\frac{k(m-1)}{2} = \frac{1}{m}Y_i - \left(ai + b + \frac{n-k}{2}a\right) \end{aligned}$$

□

Satz 16.19: Sei $t = (1, 2, \ldots, n)$ und $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, wobei $n = mk$. Wir bezeichnen das arithmetische Mittel der y_t, die zur Saison $i \in 1 \ldots k$ gehören, mit A_i, formal

$$A_i := \frac{1}{m}\sum_{j=0}^{m-1} y_{i+jk}$$

Dann suchen wir a, $b \in \mathbb{R}$ und $s = (s_1, s_2, \ldots, s_k) \in \mathbb{R}^k$ so dass wir eine möglichst gute Approximation $y_t \approx at + b + s_i$ erhalten. D.h. wir minimieren den folgenden Abstand:

$$d(a,b,s) := \sqrt{\sum_{i=1}^{k}\sum_{j=0}^{m-1}\big(a(i+jk) + b + s_i - y_{i+jk}\big)^2}$$

Ferner soll das arithmetische Mittel der s_i verschwinden $A(s) = 0$. Die Lösung dieses Problems ergibt sich zu:

$$\begin{aligned} a &= \frac{12}{n^2-k^2}A(ty) - \frac{6}{n+k}A(y) - \frac{12}{k(n^2-k^2)}\sum_{i=1}^{k} iA_i \\ b &= A(y) - \frac{n+1}{2}a \\ s_i &= A_i - A(y) + \left(\frac{k+1}{2} - i\right) a \end{aligned}$$

Beweis: Eigentlich ist b überflüssig, da der konstante Anteil ja auch in die saisonalen Einflüsse s_i integriert werden kann. Deswegen betrachten wir zunächst: $d(a,s) := d(a,0,s)$. Wie üblich (da $x \mapsto x^2$ ordungserhaltend ist) genügt es $q(a,s) := d(a,s)^2$ zu minimieren. Wir berechnen also:

$$\begin{aligned} q(a,s) &= \sum_{i=1}^{k}\sum_{j=0}^{m-1} \big(a(i+jk) + s_i - y_{i+jk}\big)^2 \\ &= \sum_{i=1}^{k}\sum_{j=0}^{m-1} \big(a^2(i+jk)^2 + s_i^2 + y_{i+jk}^2\big) \\ &\quad +2\sum_{i=1}^{k}\sum_{j=0}^{m-1} \big(as_i(i+jk) - ay_{i+jk}(i+jk) - s_i y_{i+jk}\big) \end{aligned}$$

Wir vereinfachen nun die einzelnen Terme in diesem Ausdruck. Die meisten Vereinfachungen sind offensichtlich, die nicht offensichtlichen führen wir aus:

$$\begin{aligned} \sum_{i=1}^{k}\sum_{j=0}^{m-1} a^2(i+jk)^2 &= a^2\sum_{t=1}^{n} t^2 = \frac{(2n+1)(n+1)n}{6}a^2 \\ 2\sum_{i=1}^{k}\sum_{j=0}^{m-1} as_i(i+jk) &= 2a\sum_{i=1}^{k}\left(mi + \frac{(m-1)m}{2}k\right)s_i \\ &= 2am\sum_{i=1}^{k} is_i + an(m-1)\sum_{i=1}^{k} s_i \end{aligned}$$

Setzen wir diese Ausdrücke wieder in $q(a,s)$ ein, so wird dieser Ausdruck zu

$$\begin{aligned} q(a,s) &= \frac{(2n+1)(n+1)n}{6}a^2 + m\sum_{i=1}^{k} s_i^2 + \sum_{t=1}^{n} y_t^2 + 2am\sum_{i=1}^{k} is_i \\ &\quad +an(m-1)\sum_{i=1}^{k} s_i - 2a\sum_{t=1}^{n} ty_t - 2\sum_{i=1}^{k} s_i \sum_{j=0}^{m-1} y_{i+jk} \end{aligned}$$

Damit dieser Ausdruck minimal wird, müssen die partiellen Ableitungen nach a und allen s_i verschwinden. Werten wir zunächst die Ableitung nach a aus, $\partial_a q(a,s) = 0$, erhalten wir

$$\begin{aligned}
\frac{(2n+1)(n+1)n}{3}a + 2m\sum_{i=1}^{k} is_i + n(m-1)\sum_{i=1}^{k} s_i - 2\sum_{t=1}^{n} ty_t &= 0 \\
\Longrightarrow \quad \frac{(2n+1)(n+1)}{3}a + \sum_{i=1}^{k}\left(\frac{2i}{k} + m - 1\right)s_i &= 2\sum_{t=1}^{n} ty_t
\end{aligned}$$

Nun werten wir auch die Ableitungen nach s_i aus, $\partial_{s_i} q(a,s) = 0$, aus diesen erhalten wir (für jedes $i \in 1 \ldots k$) die Gleichung

$$\begin{aligned}
2ms_i + 2ami + an(m-1) - 2\sum_{j=0}^{m-1} y_{i+jk} &= 0 \\
\Longrightarrow \quad s_i &= A_i - \left(i + \frac{n-k}{2}\right)a
\end{aligned}$$

Wir haben das lineare Gleichungssystem also bereits separiert und können diese Ausdrücke für s_i wieder in die obige Gleichung einsetzen. Dadurch erhalten wir eine Gleichung, die nur noch a entält

$$\frac{(2n+1)(n+1)}{3}a + \sum_{i=1}^{k}\left(\frac{2i}{k} + m - 1\right)\left(A_i - \left(i + \frac{n-k}{2}\right)a\right) = 2\sum_{t=1}^{n} ty_t$$

Der Übersichtlichkeit halber betrachten wir einen dieser Terme wieder isoliert. Wir ersparen uns die Details der Rechnung und erhalten

$$\begin{aligned}
& \sum_{i=1}^{k}\left(\frac{2i}{k} + m - 1\right)\left(A_i - \left(i + \frac{n-k}{2}\right)a\right) \\
= \; & \frac{2}{k}\sum_{i=1}^{k} iA_i + (m-1)kA(y) - \frac{2}{k}a\frac{(2k+1)(k+1)k}{6} \\
& -(m-1)\frac{k(k+1)}{2}a - (m-1)\frac{k(k+1)}{2}a - \frac{(n-k)^2}{2}a \\
= \; & \frac{2}{k}\sum_{i=1}^{k} iA_i + (n-k)A(y) - \frac{1}{6}\left(3n^2 + 6n + k^2 + 2\right)a
\end{aligned}$$

Diesen Ausdruck können wir schließlich in die voran gegangene Gleichung einsetzen und damit nach a auflösen. Man rechnet leicht nach, dass

$$\begin{aligned}
\frac{1}{6}\left(n^2 - k^2\right)a + \frac{2}{k}\sum_{i=1}^{k} iA_i + (n-k)A(y) &= 2\sum_{t=1}^{n} ty_t \\
\Longrightarrow \quad a &= \frac{12}{n^2-k^2}A(ty) - \frac{6}{n+k}A(y) - \frac{12}{k(n^2-k^2)}\sum_{i=1}^{k} iA_i
\end{aligned}$$

Wir haben also das a gefunden, dass unser Problem löst. Da aber auch das arithmetische Mittel der s_i verschwinden soll, fangen wir dieses in b auf:

$$\begin{aligned} b &:= \frac{1}{k}\sum_{i=1}^{k} s_i = \frac{1}{k}\sum_{i=1}^{k}\left(A_i - \left(i + \frac{n-k}{2}\right)a\right) \\ &= A(y) - \frac{1}{k}\frac{k(k+1)}{2}a - \frac{n-k}{2}a \\ &= A(y) - \frac{n+1}{2}a \end{aligned}$$

Im folgenden nehmen wir also $s_i' := s_i - b$, dann ist klar, dass das arithmetische Mittel der s_i' verschwindet: $A(s_i') = A(s_i - b) = A(s_i) - b = 0$. Damit bilden also a, b und die s_i' eine Lösung des Problems, wobei

$$s_i' = s_i - b = A_i - A(y) + \left(\frac{k+1}{2} - i\right)a$$

□

Zu Klassierungen:

Definition 16.20: Ist S eine beliebige Menge, dann heißt eine Menge $\mathcal{S} = \{\, S_i \mid i \in I \,\}$ von Teilmengen $S_i \subseteq S$ eine **Partition** oder auch **Klassierung** von S, falls gilt: (1) keine der *Klassen* S_i ist leer, d.h. für alle $i \in I$ gilt $S_i \neq \emptyset$, (2) je zwei verschiedene Teilmengen aus $\mathcal{S}$ sind *disjunkt*

$$i, j \in I \text{ mit } i \neq j \implies S_i \cap S_j = \emptyset$$

und (3) die Klassen S_i *überdecken* S, d.h. zu jedem $s \in S$ gibt es ein $i \in I$ mit $x \in S_i$. Oder anders ausgedrückt: die Vereinigung aller Teilmengen S_i aus $\mathcal{S}$ ergibt wieder die ganze Menge S, formal

$$S = \bigcup_{i \in I} S_i$$

Bemerkung 16.21: Im folgenden sei stets S eine Menge von Merkmalsträgern, und $x = (x_1, x_2, \dots, x_n)$, wobei $x_t \in S$ für $t \in 1 \dots n$. Ferner sei $\mathcal{S} = \{\, S_1, S_2, \dots, S_m \,\}$ eine Klassierung von S. Dann bezeichnen wir die *absolute* bzw. *relative Häufigkeit* der Klasse S_i mit

$$\begin{aligned} n_i &:= \#\{\, t \in 1 \dots n \mid x_t \in S_i \,\} \\ h_i &:= \frac{n_i}{n} \end{aligned}$$

Da $\mathcal{S}$ eine Klassierung ist finden wir $n_1 + n_2 + \dots + n_m = n$ oder anders ausgedrückt $h_1 + h_2 + \dots + h_m = 1$. Den bisherigen, unklassierten Fall erhalten wird durch die einelementigen Klassen $S_i := \{\, s_i \,\}$, wobei die s_i wieder die verschiedenen Werte unter den x_t sein sollen.

Beispiel 16.22: Zumeist ist $S = [a, b[$ ein halboffenes Intervall und wir führen die Klassen $S_i = [a_{i-1}, a_i[$ ein, wobei $i \in 1 \dots m$ und

$$a \;=\; a_0 \;<\; a_1 \;<\; \dots \;<\; a_m \;=\; b$$

Die Klassen S_i sind dann nach Konstruktion (als halboffene Intervalle) disjunkt und $S = S_1 \cup S_2 \cup \dots \cup S_m$ folgt daraus, dass $a_0 = a$ und $a_m = b$ sind. In diesem Fall benutzen wir wieder die *Klassenmitten* s_i^* und die *Klassenbreiten* w_i

$$\begin{aligned} s_i^* \;&:=\; \frac{a_{i-1} + a_i}{2} \\ w_i \;&:=\; a_i - a_{i-1} \end{aligned}$$

Die *Standardannahme* lautet nun, dass alle $x \in S_i$ dieselbe Häufigkeit haben, innerhalb der Klasse S_i besteht also die *Häufigkeitsdichte*

$$h_i^* \;:=\; \frac{h_i}{w_i}$$

Satz 16.23: Sei $S = [a, b[$ klassiert, durch $S_i = [a_{i-1}, a_i[$, wie in (16.22) beschrieben. Und sei $x = (x_1, x_2, \dots, x_n)$ wobei $x_t \in S$ für $t \in 1 \dots n$. Dann gilt für das *(klassierte) arithmetische Mittel* $A^*(x)$ bzw. für die *(klassierte) Varianz* $\sigma_2^*(x)$

$$\begin{aligned} A^*(x) \;&:=\; \sum_{i=1}^{m} \int_{a_{i-1}}^{a_i} h_i^* s \, ds \;=\; \sum_{i=1}^{m} h_i s_i^* \\ \sigma_2^*(x)^2 \;&:=\; \sum_{i=1}^{m} h_i^* \int_{a_{i-1}}^{a_i} (s - A^*(x))^2 \, ds \\ &=\; \sum_{i=1}^{m} h_i \big(s_i^* - A^*(x)\big)^2 + \frac{1}{12} \sum_{i=1}^{m} h_i w_i^2 \end{aligned}$$

Beweis: Bekanntermaßen ist $s^2/2$ eine Stammfunktion von s, also können wir das Intergal berechnen, durch

$$\begin{aligned} \int_{a_{i-1}}^{a_i} h_i^* s \, ds \;&=\; h_i^* \int_{a_{i-1}}^{a_i} s \, ds \;=\; h_i^* \left(\frac{1}{2} a_i^2 - \frac{1}{2} a_{i-1}^2 \right) \\ &=\; \frac{h_i}{a_i - a_{i-1}} \frac{a_i^2 - a_{i-1}^2}{2} \;=\; h_i \frac{a_i + a_{i-1}}{2} \;=\; h_i s_i^* \end{aligned}$$

Durch Summation über i (von 1 bis m) ergibt sich also die Identität für das klassierte arithmetische Mittel. Zur Abkürzung sei nun $\overline{x} := A^*(x)$. Für die Varianz müssen wir ganz analog das Integral auswerten:

$$
\begin{aligned}
h_i^* \int_{a_{i-1}}^{a_i} (s - \overline{x})^2 \, ds &= h_i^* \int_{a_{i-1}}^{a_i} (s^2 - 2s\overline{x} + \overline{x}^2) \, ds \\
&= h_i^* \left(\frac{1}{3} \left(a_i^3 - a_{i-1}^3\right) - \overline{x} \left(a_i^2 - a_{i-1}^2\right) + \overline{x}^2 (a_i - a_{i-1}) \right)
\end{aligned}
$$

Nach Definition ist $w_i = a_i - a_{i-1}$ und damit sieht man leicht, dass $a_i^2 - a_{i-1}^2 = 2 w_i s_i^*$ und $a_i^3 - a_{i-1}^3 = w_i (a_i^2 + a_{i-1} a_i + a_{i-1}^2)$. Nutzt man nun noch aus, dass $h_i^* = h_i / w_i$, so gelangt man zu der Gleichung

$$
\begin{aligned}
h_i^* \int_{a_{i-1}}^{a_i} (s - \overline{x})^2 \, ds &= h_i \left(\frac{1}{3} \left(a_i^2 + a_{i-1} a_i + a_{i-1}^2\right) - 2 \overline{x} s_i^* + \overline{x}^2 \right) \\
&= h_i \left(\frac{1}{3} \left(a_i^2 + a_{i-1} a_i + a_{i-1}^2\right) - (s_i^*)^2 + (s_i^* - \overline{x})^2 \right) \\
&= h_i \left(\frac{1}{12} a_i^2 + \frac{1}{6} a_{i-1} a_i + \frac{1}{12} a_{i-1}^2 + (s_i^* - \overline{x})^2 \right) \\
&= h_i \left(\frac{1}{12} w_i^2 + (s_i^* - \overline{x})^2 \right)
\end{aligned}
$$

Die Varianz erhält man durch Summation dieser Ausdrücke über i (von 1 bis m). Sortieren wir diese Summe entsprechend um, so finden wir also

$$
\sigma_2^*(x)^2 := \sum_{i=1}^{m} h_i \left(s_i^* - \overline{x}\right)^2 + \frac{1}{12} \sum_{i=1}^{m} h_i w_i^2
$$

□

Satz 16.24: Sei $S = [a, b[$ klassiert, durch $S_i = [a_{i-1}, a_i[$, wie in (16.22) beschrieben. Dann definieren wir die *empirische Verteilungsfunktion* H als die stückweise affine Funktion zu den Stützstellen $(a_i \mid H_i)$ (mit $i \in 1 \ldots m$). Die Funktion $H : \mathbb{R} \to [0, 1]$ ist also gegeben durch

$$
H(s) = \begin{cases} 0 & \text{für} \quad s < a_0 \\ H_{k-1} + h_k^*(s - a_{k-1}) & \text{für} \quad a_{k-1} \le s < a_k \\ 1 & \text{für} \quad s \ge a_m \end{cases}
$$

Das j-te, *(klassierte) Quartil* $Q_j^*(x)$ (wobei $j \in 1 \ldots 3$) ist definiert, als die Stelle, an der H den Wert $j/4$ annimmt. Explizit lässt sich dies berechnen, durch (wenn $k \in 1 \ldots m$ mit $H_{k-1} < j/4 \le H_k$)

$$
Q_j^*(x) := H^{-1}(j/4) = a_k - \frac{H_k - H_k}{h_k^*}
$$

Das zweite klassierte Quartil wird auch *(klassierter) Zentralwert* genannt $Z^* := Q_2^*(x)$. Und ist $k \in 1 \dots m$ mit $a_{k-1} \leq Z^*(x) \leq a_k$ so gilt für die *(klassierte), absolute, mittlere Abweichung*

$$\begin{aligned}\sigma_1^*(x) &:= \sum_{i=1}^{m} h_i^* \int_{a_{i-1}}^{a_i} |s - Z^*(x)|\, ds \\ &= \sum_{i=1}^{m} h_i |s_i^* - Z^*(x)| - h_k^* \left(\frac{w_k}{2} - |s_k^* - Z^*(x)|\right)\end{aligned}$$

Beweis: Wir beweisen zunächst die explizite Darstellung der empirischen Verteilungsfunktion. Für $s < a_0$ ist $H(s) = 0$ klar, ebenso $H(s) = 1$ für $s > a_m$. Sei also $a_{k-1} \leq s < a_k$, dann liegt $H(s)$ auf der Geraden durch die Punkte $(a_{k-1} \mid H_{k-1})$ und $(a_k \mid H_k)$. Die Steigung der Geraden ist also $(H_k - H_{k-1})/(a_k - a_{k-1}) = h_k/w_k = h_k^*$. Die Gerade und damit H in diesem Bereich ist also wie behauptet $h_k^*(s - a_{k-1}) + H_{k-1}$. Wir berechnen damit die Quartile. Nach Definition gilt

$$\begin{aligned}j/4 &= H\left(Q_j^*(x)\right) \;=\; H_{k-1} + h_k^*\left(Q_j^*(x) - a_{k-1}\right) \\ &= H_k - h_k - h_k^*\left(a_{k-1} - Q_j^*\right) \;=\; H_k - h_k^*\left(w_k + a_{k-1} - Q_j^*\right) \\ &= H_k - h_k^*\left(a_k - Q_j^*\right)\end{aligned}$$

Lösen wir diese Gleichung nach $Q_j^*(x)$ auf erhalten wir also wie behauptet $Q_j^*(x) = a_k - (H_k - j/4)/h_k^*$. Es bleibt also noch die Formel für die mittlere, absolute Abweichung zu beweisen. Zur Abkürzung setzen wir $\overline{x} := Z^*(x)$. Sei nun $k \in 1 \dots m$ so fixiert, dass $a_{k-1} \leq \overline{x} \leq a_k$. Zunächst betrachten wir aber den Fall $\overline{x} \leq a_{i-1}$, dann ist

$$\begin{aligned}h_i^* \int_{a_{i-1}}^{a_i} |s - \overline{x}|\, ds &= h_i^* \int_{a_{i-1}}^{a_i} s - \overline{x}\, ds \\ &= h_i^* \left(\frac{a_i^2 - a_{i-1}^2}{2} - \overline{x}(a_i - a_{i-1})\right) \\ &= \frac{h_i}{w_i}\left(s_i^* w_i - \overline{x} w_i\right) \;=\; h_i\left(s_i^* - \overline{x}\right)\end{aligned}$$

Ganz analog (nur mit negativem Vorzeichen) ist der Fall $\overline{x} \leq a_i$. Insgesamt erhalten wir also (in dem Fall dass $\overline{x} \notin \,]a_{i-1}, a_i[$ ist)

$$h_i^* \int_{a_{i-1}}^{a_i} |s - \overline{x}|\, ds = h_i |s_i^* - \overline{x}|$$

Es bleibt also nur noch den Fall $\overline{x} \in [a_{k-1}, a_k]$ zu betrachten. In diesem Fall zerlegen wir das Integral in zwei Teile

$$\begin{aligned}\int_{a_{k-1}}^{a_k} |s - \overline{x}|\, ds &= \int_{a_{k-1}}^{\overline{x}} \overline{x} - s\, ds + \int_{\overline{x}}^{a_k} s - \overline{x}\, ds \\ &= \overline{x}(\overline{x} - a_{k-1}) - \int_{a_{k-1}}^{\overline{x}} s\, ds + \int_{\overline{x}}^{a_k} s\, ds - \overline{x}(a_k - \overline{x})\end{aligned}$$

$$
\begin{aligned}
&= 2\overline{x}^2 - \overline{x}(a_{k-1} + a_k) - \frac{\overline{x}^2 - a_{k-1}^2}{2} + \frac{a_k^2 - \overline{x}^2}{2} \\
&= \overline{x}^2 - 2\overline{x}s_k^* + s_k^* - s_k^* + \frac{a_{k-1}^2 + a_k^2}{2} \\
&= \left(\overline{x} - s_k^*\right)^2 - s_k^* + \frac{a_{k-1}^2 + a_k^2}{2} \\
&= \left(\overline{x} - s_k^*\right)^2 + \frac{1}{4}w_k^2
\end{aligned}
$$

Die letzte Gleichung $-s_k^* + (a_{k-1}^2 + a_k^2)/2 = w_k^2/4$ lässt sich leicht nachrechnen, ist aber nicht offensichtlich. Schließlich haben wir alle Fälle behandelt und können die absolute, mittlere Abweichung auswerten:

$$
\begin{aligned}
\sigma_1^*(x) &= \sum_{i=1}^{m} h_i^* \int_{a_{i-1}}^{a_i} |s - \overline{x}|\, ds \\
&= \sum_{i \neq k} h_i^* \int_{a_{i-1}}^{a_i} |s - \overline{x}|\, ds + h_k^* \int_{a_{k-1}}^{a_k} |s - \overline{x}|\, ds \\
&= \sum_{i \neq k} h_i |s_i^* - \overline{x}| + \left(\overline{x} - s_k^*\right)^2 + \frac{1}{4}w_k^2 \\
&= \sum_{i=1}^{m} h_i |s_i^* - \overline{x}| - h_k |s_k^* - \overline{x}| + \left(\overline{x} - s_k^*\right)^2 + \frac{1}{4}w_k^2 \\
&= \sum_{i=1}^{m} h_i |s_i^* - \overline{x}| + \left(\frac{w_k}{2}\right)^2 - h_k |s_k^* - \overline{x}| + |\overline{x} - s_k^*|^2 \\
&= \sum_{i=1}^{m} h_i |s_i^* - \overline{x}| + \left(\frac{w_k}{2} - |\overline{x} - s_k^*|\right)^2
\end{aligned}
$$

□

Satz 16.25:
Sie $S = [a, b[$ klassiert, durch $S_i = [a_{i-1}, a_i[$, wie in (16.22) beschrieben. Die Lorenzkurve entstand (im unkalssierten Fall) durch Verbinden der Stützstellen $(H_k \mid L_k)$. Da aber die kumulierte Häufigkeit im klassierten Fall aber stetig wächst (siehe H in (16.24)), muss man zu einer Kurve der Art $(H(s) \mid L(s))$ übergehen. Dabei muss L wieder die relativen, kumulierten Merkmalssummen angeben. Aufgrund der Standardannahme lautet die passende Übersetzung (für $a_{k-1} \leq s \leq a_k$)

$$
\begin{aligned}
X &:= nA^*(x) = \sum_{i=1}^{m} n_i s_i^* \\
L(s) &:= \frac{1}{X}\left(\sum_{i=1}^{k-1} n_i s_i^* + \int_{a_{k-1}}^{s} n_k \frac{r}{w_k}\, dr\right) \\
&= \frac{1}{A^*(x)}\left(\sum_{i=1}^{k-1} h_i s_i^* + h_k^* \frac{s^2 - a_{k-1}^2}{2}\right)
\end{aligned}
$$

Als Funktion $L^* : [0,1] \to [0,1]$ ist die klassierte Lorenzkurve also gegeben, durch $L^* := LH^{-1}$. Konkret bedeutet dass (für $H_{k-1} \leq u \leq H_k$)

$$L^*(u) \quad = \quad L(s) \quad \text{wobei} \quad s \;=\; a_{k-1} + \frac{u - H_{k-1}}{h_k^*}$$

Damit ist L^* dann eine stetige, stückweise quadratische Funktion, genauer lässt sich L^* (für $H_{k-1} \leq u \leq H_k$) berechnen, als $L^*(u) =$

$$\frac{1}{A^*(x)} \left(\frac{1}{2h_k^*} u^2 + \left(a_{k-1} - \frac{H_{k-1}}{h_k^*} \right) u + \sum_{i=1}^{k-1} h_i s_i^* + \frac{H_{k-1}^2}{2h_k^*} - a_{k-1} H_{k-1} \right)$$

Der Gini-Koeffizient ist wieder die doppelte Fläche zwischen der Diagonalen und der Lorenzkurve. Formal also

$$\begin{aligned} R^* \quad &:= \quad 2 \int_0^1 u - L^*(u)\, du \quad = \quad 1 - 2 \int_0^1 L^*(u)\, du \\ &= \quad 1 - \sum_{i=1}^m h_i (L_{i-1}^* + L_i^*) + \frac{1}{6A^*(x)} \sum_{i=1}^m h_i^2 w_i \end{aligned}$$

Beweis:
Wir rechnen zunächst die angegebene Formel für die kumulierten, relativen Merkmalssummen nach. Sei also $a_{k-1} \leq s \leq a_k$, dann ist

$$\begin{aligned} L(s) \quad &:= \quad \frac{1}{X} \left(\sum_{i=1}^{k-1} n_i s_i^* + \int_{a_{k-1}}^{s} n_k \frac{r}{w_k}\, dr \right) \\ &= \quad \frac{1}{nA^*(x)} \left(\sum_{i=1}^{k-1} n h_i s_i^* + n \frac{h_k}{w_k} \int_{a_{k-1}}^{s} r\, dr \right) \\ &= \quad \frac{1}{A^*(x)} \left(\sum_{i=1}^{k-1} h_i s_i^* + h_k^* \int_{a_{k-1}}^{s} r\, dr \right) \\ &= \quad \frac{1}{A^*(x)} \left(\sum_{i=1}^{k-1} h_i s_i^* + h_k^* \frac{s^2 - a_{k-1}^2}{2} \right) \end{aligned}$$

Für die angegebene Formel der klassierten Lorenzkurve muss man sich ein bisschen mehr anstrengen: nach Definition ist $L^*(u) = L(s)$ wobei $s = H^{-1}(u) = (u - H_{k-1})/h_k^* + a_{k-1}$. Wir berechnen also vorab

$$\begin{aligned} s^2 &= \left(\frac{u - H_{k-1}}{h_k^*} + a_{k-1} \right)^2 \\ &= \left(\frac{u - H_{k-1}}{h_k^*} \right)^2 + \frac{2a_{k-1}}{h_k^*}(u - H_{k-1}) + a_{k-1}^2 \\ h_k^* \frac{s^2 - a_{k-1}^2}{2} &= \frac{(u - H_{k-1})^2}{2h_k^*} + a_{k-1}(u - H_{k-1}) \\ &= \frac{u^2 - 2H_{k-1}u + H_{k-1}^2}{2h_k^*} + a_{k-1}(u - H_{k-1}) \end{aligned}$$

$$= \frac{1}{2h_k^*}u^2 + \left(a_{k-1} - \frac{H_{k-1}}{h_k^*}\right)u + \frac{H_{k-1}^2}{2h_k^*} - a_{k-1}H_{k-1}$$

Aufgrund der oben bewiesenen Formel für $L(s)$ und der soeben bewiesenen Formel setzt sich die klassierte Lorenzkurve also zusammen, zu $L^*(u) =$

$$\frac{1}{A^*(x)}\left(\frac{1}{2h_k^*}u^2 + \left(a_{k-1} - \frac{H_{k-1}}{h_k^*}\right)u + \sum_{i=1}^{k-1} h_i s_i^* + \frac{H_{k-1}^2}{2h_k^*} - a_{k-1}H_{k-1}\right)$$

Wir werten also den Gini-Koeffizienten aus. Für diesen haben wir 2 weitere Formeln angegeben, von denen die erste unmittelbar einsichtig ist:

$$\begin{aligned} R^* &= 2\int_0^1 u - L^*(u)\,du &= 2\int_0^1 u\,du - 2\int_0^1 L^*(u)\,du \\ &= 2\frac{1^2 - 0^2}{2} - 2\int_0^1 L^*(u)\,du &= 1 - 2\int_0^1 L^*(u)\,du \end{aligned}$$

Nach dieser Formel genügt es also das Intergal über die klassierte Lorenzkurve auszuwerten. Da wir die Lorenzkurve auch schon explizit berechnet haben, ist dies auch mit etwas Konzentration möglich:

$$\begin{aligned} \int_0^1 L^*(u)\,du &= \sum_{k=1}^m \int_{H_{k-1}}^{H_k} L^*(u)\,du \\ &= \frac{1}{A^*(x)}\sum_{k=1}^m \int_{H_{k-1}}^{H_k} \frac{1}{2h_k^*}u^2 + b_k u + S_{k-1} + c_k\,du \\ &= \frac{1}{A^*(x)}\sum_{k=1}^m \left(\frac{H_k^3 - H_{k-1}^3}{6h_k^*} + b_k\frac{H_k^2 - H_{k-1}^2}{2} + c_k h_k\right) \end{aligned}$$

Dabei haben wir die folgenden Abkürzungen verwendet: $b_k = a_{k-1} - H_{k-1}/h_k^*$ und $c_k = S_{k-1} + H_{k-1}^2/(2h_k^*) - a_{k-1}H_{k-1}$, wobei

$$S_{k-1} := \sum_{i=1}^{k-1} h_i s_i^*$$

Man beachte dass daher $S_k = S_{k-1} + h_k s_k^*$ und $S_m = A^*(x)$ gilt. Nun ist aber $H_k^2 - H_{k-1}^2 = (H_k + H_{k-1})(H_k - H_{k-1}) = (H_k + H_{k-1})h_k$ und damit

$$\begin{aligned} b_k\frac{H_k^2 - H_{k-1}^2}{2} &= \left(a_{k-1} - \frac{H_{k-1}}{h_k^*}\right)(H_k + H_{k-1})h_k \\ &= \frac{1}{2}\Big(a_{k-1}h_k - H_{k-1}w_k\Big)\Big(H_k + H_{k-1}\Big) \\ c_k h_k &= \left(S_{k-1} + \frac{H_{k-1}^2}{2h_k^*} - a_{k-1}H_{k-1}\right)h_k \\ &= S_{k-1}h_k + \frac{1}{2}H_{k-1}^2 w_k - a_{k-1}H_{k-1}h_k \end{aligned}$$

Setzt man diesen beiden Gleichungen zusammen, so erhält man durch elementare Rechung eine weitere Gleichung, die wir später verwenden werden

$$b_k \frac{H_k^2 - H_{k-1}^2}{2} + c_k h_k \quad = \quad \frac{1}{2} a_{k-1} h_k^2 - \frac{1}{2} H_{k-1} H_k w_k + S_{k-1} h_k$$

Doch zunächst verwenden wir noch, dass $H_k^3 - H_{k-1}^3 = (H_k^2 + H_k H_{k-1} + H_{k-1}^2)(H_k - H_{k-1}) = (H_k^2 + H_k H_{k-1} + H_{k-1}^2) h_k$ ist. Dies liefert

$$\frac{H_k^3 - H_{k-1}^3}{6h_k^*} \quad = \quad \frac{w_k}{6} \Big(H_k^2 + H_k H_{k-1} + H_{k-1}^2 \Big)$$

Setzen wir all diese Formeln - für das Integral der Lorenzkurve und die einzelnen Summanden die darin vorkommen - zusammen, so finden wir

$$\begin{aligned}
\int_0^1 L^*(u)\,du &= \frac{1}{A^*(x)} \sum_{k=1}^m \left(\frac{H_k^3 - H_{k-1}^3}{6h_k^*} + b_k \frac{H_k^2 - H_{k-1}^2}{2} + c_k h_k \right) \\
&= \frac{1}{A^*(x)} \sum_{k=1}^m \left(\frac{w_k}{6} \Big(H_k^2 + H_k H_{k-1} + H_{k-1}^2 \Big) \right. \\
&\quad \left. + \frac{1}{2} a_{k-1} h_k^2 - \frac{1}{2} H_{k-1} H_k w_k + S_{k-1} h_k \right) \\
&= \frac{1}{A^*(x)} \sum_{k=1}^m \left(\frac{w_k}{6} \Big(H_k^2 - 2H_k H_{k-1} + H_{k-1}^2 \Big) \right. \\
&\quad \left. + \frac{1}{2} a_{k-1} h_k^2 + S_{k-1} h_k \right) \\
&= \frac{1}{A^*(x)} \sum_{k=1}^m \left(\frac{w_k}{6} h_k^2 + \frac{1}{2} a_{k-1} h_k^2 + S_{k-1} h_k \right) \\
&= \frac{1}{A^*(x)} \sum_{k=1}^m h_k \left(\frac{h_k w_k}{6} + \frac{h_k a_{k-1}}{2} - h_k s_k^* + S_k \right) \\
&= \frac{1}{A^*(x)} \sum_{k=1}^m h_k \left(\frac{h_k w_k}{6} - \frac{h_k a_k}{2} + S_k \right) \\
&= \frac{1}{A^*(x)} \sum_{k=1}^m \frac{h_k}{2} \left(2S_k - h_k a_k + \frac{h_k w_k}{3} \right) \\
&= \frac{1}{A^*(x)} \sum_{k=1}^m \frac{h_k}{2} \left(2S_{k-1} + 2h_k s_k^* - h_k a_k + \frac{h_k w_k}{3} \right)
\end{aligned}$$

An dieser Stelle verwenden wir nun $h_k s_k^* - h_k a_k = h_k(a_{k-1}/2 + a_k/2 - a_k) = h_k(a_{k-1}/2 - a_k/2) = -h_k w_k/2$. Setzen wir dies ein, so folgt

$$\begin{aligned}
\int_0^1 L^*(u)\,du &= \frac{1}{A^*(x)} \sum_{k=1}^m \frac{h_k}{2} \left(2S_{k-1} + h_k s_k^* - \frac{h_k w_k}{2} + \frac{h_k w_k}{3} \right) \\
&= \frac{1}{A^*(x)} \sum_{k=1}^m \frac{h_k}{2} \left(2S_{k-1} + h_k s_k^* - \frac{h_k w_k}{6} \right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{k=1}^{m}\frac{h_k}{A^*(x)}\left(2S_{k-1}+h_k s_k^*\right) - \frac{1}{12A^*(x)}\sum_{k=1}^{m} h_k^2 w_k \\
&= \frac{1}{2}\sum_{k=1}^{m}\frac{h_k}{A^*(x)}\left(S_{k-1}+S_k\right) - \frac{1}{12A^*(x)}\sum_{k=1}^{m} h_k^2 w_k \\
&= \frac{1}{2}\sum_{k=1}^{m} h_k\left(L_{k-1}^*+L_k^*\right) - \frac{1}{12A^*(x)}\sum_{k=1}^{m} h_k^2 w_k \\
R^* &= 1-2\int_0^1 L^*(u)\,du \\
&= 1-\sum_{k=1}^{m} h_k\left(L_{k-1}^*+L_k^*\right) + \frac{1}{6A^*(x)}\sum_{k=1}^{m} h_k^2 w_k
\end{aligned}
$$

□

Zur Wahrscheinlichkeitsrechnung:

Die allgemeine Behandlung der Wahrscheinlichkeitstheorie stößt auf ein paar unerwartete Widerstände - vor allem das Problem, dass es unmöglich ist *jeder* Teilmenge A von $S = \mathbb{R}$ eine sinnvolle Wahrscheinlichkeit $P(A)$ zuzuordnen. Das ist zwar auch nicht nötig (denn nicht alle Teilmengen kommen als Ereignismengen in Frage) führt aber zu einigen Komplikationen, auf die an dieser Stelle eingegangen werden soll. In Kapitel 12 haben wir die mathematische Exaktheit daher ein wenig vernachlässigt, dies soll nun aber nachgeholt werden.

Definition 16.26:
Sei S eine beliebige Menge. Dann nennen wir eine Kollektion $\Omega \subseteq \mathcal{P}(S)$ von Teilmengen von S einen **Mengenring** auf S, falls folgendes gilt:

(1) $S \in \Omega$

(2) $A \in \Omega \implies S \setminus A \in \Omega$

(3) $A, B \in \Omega \implies A \cup B \in \Omega$

Da wir den Schnitt zweier Teilmengen A und $B \subseteq S$ auch ausdrücken können, als $A \cap B = S \setminus ((S \setminus A) \cup (S \setminus B))$ und die Differenz $B \setminus A = B \cap (S \setminus A)$ ist, erhalten wir dann auch gleich die weiteren Eigenschaften

(4) $A, B \in \Omega \implies A \cap B \in \Omega$

(5) $A, B \in \Omega \implies B \setminus A \in \Omega$

Beispiel 16.27:
Offensichtlich ist $\Omega = \mathcal{P}(S)$ selbst ein Mengenring. Dies ist auch der Fall, den wir in Kapitel 12 vorausgesetzt haben. Für $S = \mathbb{R}$ müssten wir aber einen anderen Mengenring verwenden, nämlich den Ring aller endlichen Vereinigung von Intervallen:

$$\Omega \;:=\; \{\, I_1 \cup I_2 \cup \cdots \cup I_n \mid n \in \mathbb{N},\; I_k \text{ Intervall} \,\}$$

Zur Erinnerung: ein Intervall ist eine Menge der Form $I = \{\, s \in \mathbb{R} \mid a \Box s \Box b \,\}$, wobei a und $b \in \mathbb{R}$ oder $\pm\infty$ sind und $\Box$ für die Relationen $\leq$ oder $<$ steht.

Definition 16.28:
Sei S eine beliebige Menge, und $\Omega \subseteq \mathcal{P}(S)$ ein Mengenring auf S. Dann nennen wir eine Abbildung P der Form $P : \Omega \to \mathbb{R}$, die den Teilmengen $A \subseteq S$ (mit $A \in \Omega$) ein Maß $P(A) \in \mathbb{R}$ zuordnet, eine *Wahrscheinlichkeitsfunktion*, falls für alle A, $B \in \Omega$ gilt:

(1) $P(S) \;=\; 1$

(2) $P(A) \;\geq\; 0$

(3) $A \cap B = \emptyset \quad\Longrightarrow\quad P(A \cup B) = P(A) + P(B)$

Ist nun $P : \Omega \to \mathbb{R}$ eine Wahrscheinlichkeitsfunktion auf S. Dann nennen wir P **stetig**, falls für alle absteigenden Folgen von Ereignismengen [d.h. für alle Folgen $(A_n) \subseteq \Omega$ mit $A_0 \supseteq A_1 \supseteq A_2 \supseteq \ldots$] gilt:

$$\bigcap_{n \in \mathbb{N}} A_n = \emptyset \quad\Longrightarrow\quad \lim_{n \to \infty} P(A_n) = 0$$

Satz 16.29:
Ist $P : \mathcal{P}(S) \to \mathbb{R}$ eine Wahrscheinlichkeitsfunktion auf der Menge S, dann gelten weiterhin die folgenden Aussagen für alle Teilmengen A, $B \subseteq S$:

(4) $P(\emptyset) \;=\; 0$

(5) $P(\overline{A}) \;=\; 1 - P(A)$

(5) $P(B \setminus A) \;=\; P(B) - P(A \cap B)$

(6) $P(A \cup B) \;=\; P(A) + P(B) - P(A \cap B)$

(7) $A \subseteq B \quad\Longrightarrow\quad P(A) \leq P(B)$

Beweis:

(5) Da A und $\overline{A}$ disjunkt sind (d.h. $A \cap \overline{A} = \emptyset$), erhalten aus Eigenschaft (3) $1 = P(S) = P(A \cup \overline{A}) = P(A) + P(\overline{A})$. Umgeformt also die Behauptung $P(\overline{A}) = 1 - P(A)$.

(4) Aus (5) erhalten wir insbesondere $P(\emptyset) = P(\overline{S}) = 1 - P(S) = 1 - 1 = 0$.

(6) Da A und $\overline{A}$ ganz S zerlegen, wird insbesondere auch B zerlegt, in $B = B \cap S = B \cap (A \cup \overline{A}) = (B \cap A) \cup (B \cap \overline{A})$. Mit Eigenschaft (3) also wieder $P(B) = P(B \cap A) + P(B \cap \overline{A})$. Nun ist aber $B \cap \overline{A} = B \setminus A$, also $P(B) = P(A \cap B) + P(B \setminus A)$ und damit $P(B \setminus A) = P(B) - P(A \cap B)$.

(7) Wir zerlegen die Vereinugung $A \cup B$ in drei disjunkte Teile, nämlich: $A \cup B = (A \cap B) \cup (A \setminus B) \cup (B \setminus A)$. Mit Eigenschaft (3) ist also $P(A \cup B) = P(A \cap B) + P(A \setminus B) + P(B \setminus A)$. Nach (5) können wir dies aber umschreiben, zu $P(A \cup B) = P(A \cap B) + P(A) - P(A \cap B) + P(B) - P(A \cap B) = P(A) + P(B) - P(A \cap B)$.

(8) Wegen $A \subseteq B$ können wir B disjunkt zerlegen, in $B = A \cup (B \setminus A)$. Mit Eigenschaft (3) erhalten wir also $P(B) = P(A) + P(B \setminus A)$. Nach Eigenschaft (2) ist $P(B \setminus A) \geq 0$, also $P(B) - P(A) = P(B \setminus A) \geq 0$ und damit $P(B) \geq P(A)$.

□

Definition 16.30:
Sei Ω ein Mengenring auf S und $P : \Omega \to [0,1]$ eine Wahrscheinlichkeitsfunktion. Ist ferner $X : S \to \mathbb{R}$ und $x \in \mathbb{R}$, dann bezeichnen wir die Mengen

$$\begin{aligned} \{ X = x \} &:= \{ s \in S \mid X(s) = x \} \\ \{ X \leq x \} &:= \{ s \in S \mid X(s) \leq x \} \end{aligned}$$

Dann heißt die Funktion X eine **Zufallsvariable** unter P, falls diese Mengen immer mögliche Ereignismengen von P sind. D.h. falls für alle $x \in \mathbb{R}$ gilt, dass $\{ X = x \}$ und $\{ X \leq x \} \subseteq \Omega$. Und in diesem Fall setzen wir

$$\begin{aligned} P(X = x) &:= P(\{ s \in S \mid X(s) = x \}) \\ P(X \leq x) &:= P(\{ s \in S \mid X(s) \leq x \}) \end{aligned}$$

Satz 16.31:
Sei $P : \Omega \to [0,1]$ eine stetige Wahrscheinlichkeitsfunktion auf S und ist $X : S \to \mathbb{R}$ eine Zufallsvariable unter P. Dann gilt für alle $x \in \mathbb{R}$

$$\lim_{n \to \infty} P\left(X \leq x - \frac{1}{n} \right) = P(X \leq x) - P(X = x)$$

Beweis:
Wir bezeichnen analog $\{ X < x \} := \{ s \in S \mid X(s) < x \}$. Dann erhalten wir offensichtlich eine disjunkte Vereinigung $\{ X \leq x \} = \{ X < x \} \cup \{ X = x \}$. Damit ist also $P(X \leq x) = P(X < x) + P(X = x)$. Wir betrachten nun

$$\begin{aligned} & P(X \leq x) - P\left(X \leq x - \frac{1}{n} \right) \\ = \; & P(X = x) + P(X < x) - P\left(X \leq x - \frac{1}{n} \right) \\ = \; & P(X = x) + P\left(\{ X < x \} \setminus \left\{ X \leq x - \frac{1}{n} \right\} \right) \\ = \; & P(X = x) + P\left(\left\{ s \in S \mid x - \frac{1}{n} < X(s) < x \right\} \right) \end{aligned}$$

Nun ist die Folge von Mengen $A_n := \{ s \in S \mid x - 1/n < X(s) < x \}$ aber absteigend und der Schnitt über alle A_n ist leer [denn wäre $s \in A_n$ für alle $n \geq 1$, so wäre $x - 1/n < X(s) < x$ und damit $0 < X(s) - x < 1/n$. Da $(1/n)$ aber eine Nullfolge ist, gibt es keine Zahl $X(s) - x \in \mathbb{R}$, die das erfüllen könnte]. Da P stetig ist folgt nunmehr $P(A_n) \to 0$ und damit

$$P(X \leq x) - P\left(X \leq x - \frac{1}{n}\right) \quad \to \quad P(X = x) + 0$$

$\square$

Kapitel 17

Open Publication License

I. Copyright
The copyright to each Open Publication is owned by its author(s) or designee.

II. Scope of License
The following license terms apply to all Open Publication works, unless otherwise explicitly stated in the document. Mere aggregation of Open Publication works or a portion of an Open Publication work with other works or programs on the same media shall not cause this license to apply to those other works. The aggregate work shall contain a notice specifying the inclusion of the Open Publication material and appropriate copyright notice.

Severability: If any part of this license is found to be unenforceable in any jurisdiction, the remaining portions of the license remain in force.

No Warranty: Open Publication works are licensed and provided äs is"without warranty of any kind, express or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose or a warranty of non-infringement.

III. Requirements on both unmodified and modified Versions
Any publication in standard (paper) book form shall require the citation of the original publisher and author. The publisher and author's names shall appear on all outer surfaces of the book. On all outer surfaces of the book the original publisher's name shall be as large as the title of the work and cited as possessive with respect to the title.

IV. Requirements on Modified Works
All modified versions of documents covered by this license, including translations, anthologies, compilations and partial documents, must meet the following requirements:

- The modified version must be labeled as such.
- The person making the modifications must be identified and the modifications dated.
- Acknowledgement of the original author and publisher if applicable must be retained according to normal academic citation practices. The location of the original unmodified document must be identified.
- The original author's (or authors') name(s) may not be used to assert or imply endorsement of the resulting document without the original author's (or authors') permission.

Für Anmerkungen, Hinweise und Korrekturen bin ich immer dankbar.
Sie können mich per Mail kontaktieren, oder besuchen Sie meine Homepage:

abzeidler@gmx.de, bzw.

https://my.cloudme.com/#zeidlerweb